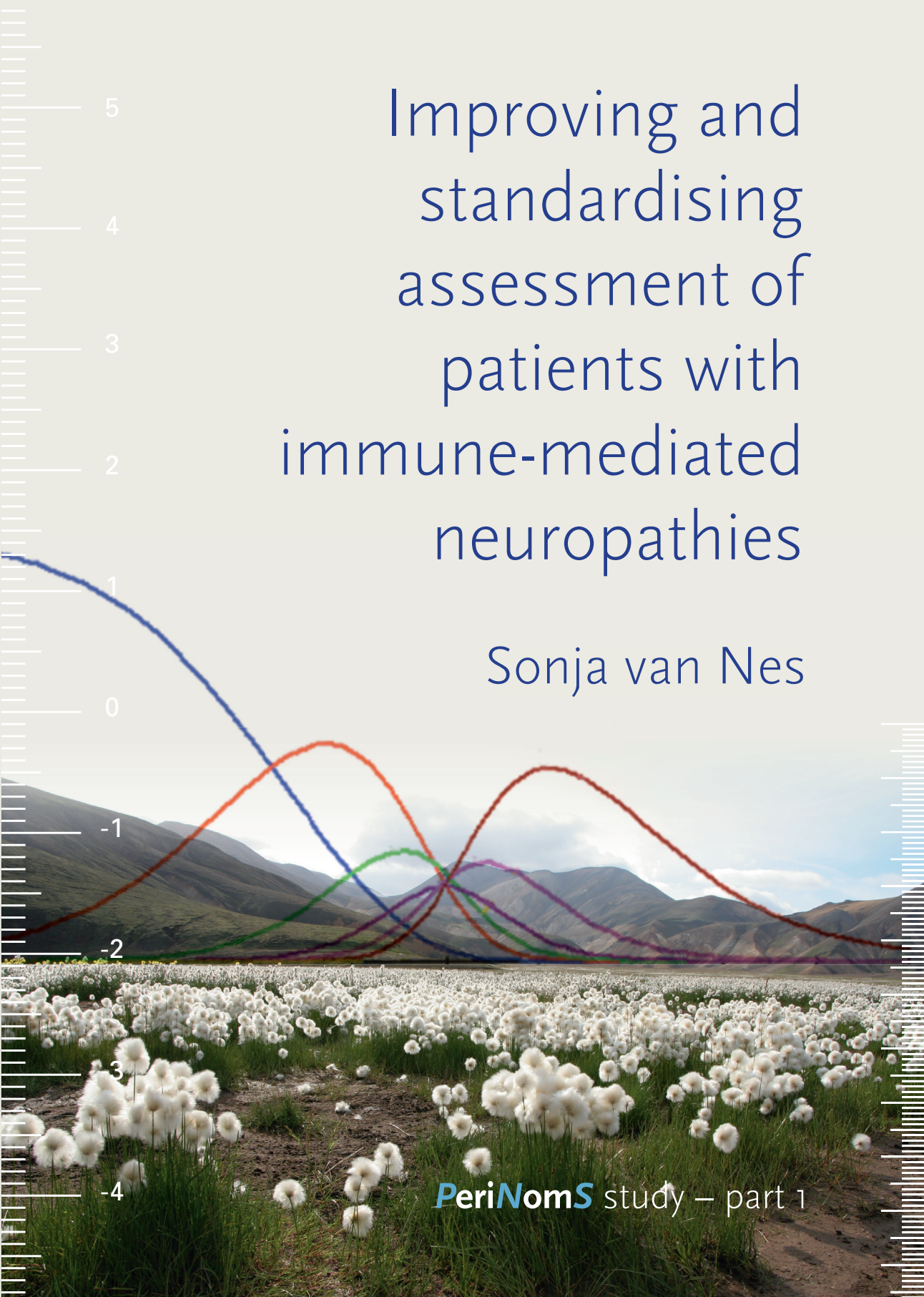
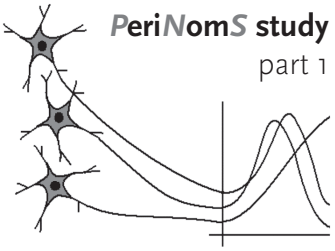
Improving and standardising assessment of patients with immune-mediated neuropathies

Sonja van Nes



PeriNomS study – part 1

Improving and standardising assessment of patients with immune-mediated neuropathies



This research project was financially supported by unrestricted grants obtained from the Talents programm by Talecris and the GBS CIDP foundation international.

ISBN 978-90-9026462-2

Lay-out & cover: Jos Hendrix, Groningen
Cover photo: Sonja van Nes; near Landmannalaugar, Iceland
Printed by Scholma Druk, Bedum

Improving and standardising assessment of patients with immune-mediated neuropathies

*P*eripheral *N*europathy outcome measures *S*tandardisation study
(*P*eri*N*om*S* study – part 1)

Het verbeteren en standaardiseren van het onderzoek van patiënten met een immuun-gemedieerde neuropathie
*P*erifere *N*europathie uitkomstmaten *S*tandaardisatie studie
(*P*eri*N*om*S* studie – deel 1)

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. dr. H.G. Schmidt
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 11 januari 2012 om 11:30 uur

door

Sonja Ingrid van Nes

Geboren te Dordrecht

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof. dr. P.A. van Doorn
Overige leden:	Prof. dr. R.A.C. Hughes Prof. dr. P.Y.K. van den Bergh Prof. dr. H.J. Stam
Copromotoren:	Dr. C.G. Faber Dr. I.S.J. Merkies

Never measure the height of a mountain
until you have reached the top.
Then you will see how low it was.

Dag Hammarskjöld, Swedish diplomat

The studies described in this thesis were performed at the Departments of Neurology of the Erasmus Medical Centre, University Hospital Rotterdam, Rotterdam, Maastricht University Medical Centre, Maastricht, Spaarne Hospital, Hoofddorp, the Netherlands in close collarobation with other national and international members of the *P*eri*N*om*S* study group from the Netherlands, UK, Belgium, Germany, France, Italy, Spain, Greece, Brazil, USA, and Canada.

Contents

Chapter 1

General introduction

Chapter 1.1

Clinical aspects of immune-mediated neuropathies

Introduction

In this thesis, the outline and the results of the first part of the *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study are described. The ***P*eri*N*om*S*** study aims to improve and standardise the assessment of patients with immune-mediated neuropathies. These disorders are potentially treatable with immuno-modulating agents, therefore, proper outcome measures are needed to detect clinically important improvement or deterioration over time.

Immune-mediated neuropathies include Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyneuropathy (CIDP), polyneuropathy associated with monoclonal gammopathy of undetermined significance (MGUSP), and multifocal motor neuropathy (MMN). Electrophysiological examination in these patients generally reveals features of a demyelinating polyneuropathy, although a subgroup of GBS patients may have predominantly axonal features (acute motor axonal neuropathy (AMAN)). Diagnostic criteria for all these neuropathies have been formulated.[1-4] From a clinical, electrophysiological, and immunological point of view there is increasing evidence that these illnesses represent part of a continuum, mainly separated by their extent of neuromuscular dysfunction, the evolution of weakness over time, and their response to treatment (table 1). In the following, a brief overview of these neuropathies is given with particular emphasis on their clinical presentation.

Guillain-Barré syndrome (GBS)

The diagnosis of GBS is generally not difficult to establish.[4, 5] Hallmark is the combination of rapidly progressive symmetrical weakness in arms and legs with or without sensory disturbances and decreased or absent tendon reflexes. By definition, maximum weakness is reached within 4 weeks, but often already within 2 weeks.[6, 7] Cranial nerve involvement, especially bilateral weakness of facial muscles, autonomic dysfunction and pain are other clinical features supporting the diagnosis. GBS is most often preceded by an infection (diarrhoea or a respiratory infection). Diagnostic criteria for GBS were originally developed to aid epidemiological field studies.[7, 8] Recently, the Brighton group suggested new criteria, however, their primary aim was proper case definition to analyse immunisation safety data.[9] GBS is a heterogeneous disease with a monophasic course of disease; some patients are only mildly affected and recover quickly, whereas others are completely paralysed within 24 hours and remain bedbound for months. About 25% of GBS patients admitted to the hospital need mechanical ventilation mainly due to weakness of the respiratory muscles. Despite treatment 3-10% of patients die, 20% are unable to walk after six months and between 25 and 85% still have residual signs and

symptoms after 2-7 years.[10] Weakness has demonstrated to have the strongest impact on ability limitations and participation restrictions.[11] Cerebrospinal fluid (CSF) examination typically shows increased protein levels (often not in the first week, but in about 90% in the second week after onset of disease). The most important reason however to perform CSF analysis is to investigate whether the cell count is normal (as it generally is in GBS), or whether it is increased (which should suspect for example neuroborreliosis). With electromyography (EMG) it is possible to characterise the pattern of GBS as either more demyelinating (acute inflammatory demyelinating polyradiculoneuropathy (AIDP)) or axonal (AMAN).[12] Patients with an intermediate 'subacute' time course have also been reported as having a subacute inflammatory demyelinating polyneuropathy (SAIDP) reaching their nadir between 4 to 8 weeks of onset.[13] About 10% of patients with GBS have 'treatment-related fluctuations', a fluctuating course after initial improvement or stabilisation after treatment with intravenous immunoglobulins (IVIg). About 5% of patients initially diagnosed as GBS turn out to progress to CIDP, acute onset CIDP (A-CIDP).[14] Beneficial effects of IVIg as well as plasma exchange have been demonstrated for GBS patients.[15-17] In many centres IVIg is the preferred treatment because of its greater convenience, availability and side-effect profile.

Chronic inflammatory demyelinating polyneuropathy (CIDP)

Patients with CIDP usually have symmetrical distal and proximal weakness that generally predominates over sensory deficit. Sensory-motor pattern is most commonly seen, although pure motor or pure sensory patterns have been reported.[1, 18, 19]

Areflexia is common. CIDP patients usually have a chronic onset of a progressive or relapsing phase over a period of more than 2 months.[20, 21] Thereafter weakness may progress or may be stable during months or years, or the patients may improve spontaneously followed by a relapsing-remitting course.[22] In general, motor impairment has a stronger impact on daily and social functioning than sensory deficit.[11]

Up to 16% of all patients with CIDP however do have an acute onset (A-CIDP), initially resembling the course of GBS. A-CIDP instead of GBS should be considered if a patient deteriorates again after 8 weeks from onset or if ≥3 treatment-related fluctuations occur.[14] Compared to GBS, the diagnosis of CIDP, in general is much more difficult to make. The American Academy of Neurology (AAN) research criteria for CIDP are rather restrictive, many patients diagnosed with CIDP by clinicians do not meet these research criteria.[20] The European Federation of Neurological Societies/Peripheral Nerve Society (EFNS/PNS) criteria including clinical and electrodiagnostic criteria were designed to overcome this limitation.[21] Prior to making the diagnosis of CIDP, it is essential to rule out other causes of chronic

polyneuropathies (or make them at least unlikely), such as hereditary neuropathies, vasculitis, cryoglobulinemia and multiple myeloma. Dyck and associates stated that the clinical diagnosis of CIDP could only be made in the absence of a systemic disease.[23] However, it seems that CIDP can also occur in the setting of some concurrent diseases. Associations with diabetes, hepatitis C, HIV infection and AIDS, lymphoma, organ transplant, connective tissue disorders, melanoma or monoclonal gammopathy of undetermined significance (MGUS) have been mentioned.[24] It is still a matter of debate whether the diagnosis of CIDP can be made in the presence of MGUS. In general it is accepted to diagnose a patient with CIDP in the presence of an IgA or IgG monoclonal paraprotein but not in the presence of an IgM paraprotein with antibodies to myelin-associated glycoprotein. Electrophysiological tests are mandatory in the EFNS/PNS criteria for CIDP and include evidence for demyelination with features such as conduction blocks, dispersion of the compound muscle action potential, increased distal latencies, or slowed conduction velocities.[25] Cerebrospinal fluid generally shows an increased protein level without a cellular reaction. IVIg, corticosteroids and treatment with plasma exchange are proven effective treatments and exert short-term or long-term clinical improvement in two-thirds of patients.[26-28] A recent 'negative' trial with methotrexate (MTX) raised the question whether some included patients might have had inactive disease and, therefore, were unlikely to respond to a new therapy like MTX beforehand.[29] This stresses the need to periodically taper or discontinue maintenance treatment for CIDP to avoid over-treatment. Especially in these patients, proper outcome measures would be very helpful to determine subtle clinically relevant changes over time.

Polyneuropathy associated with monoclonal gammopathy of undetermined significance (MGUSP)

A clone of plasma cells in the bone marrow may produce a monoclonal gammopathy that can be detected by investigating peripheral blood. This proliferative process can be low grade as in monoclonal gammopathy of undetermined significance (MGUS) or it can be 'malignant' as in multiple myeloma or M. Waldenström. IgA and IgG monoclonal gammopathies are associated with idiopathic CIDP. In general, patients with MGUS-related CIDP are somewhat older at onset of symptoms and on average have more frequent sensory loss with less severe weakness, despite similar motor conduction findings when compared to CIDP patients without MGUS. About half of the patients with polyneuropathy and an IgM monoclonal gammopathy have antibodies against myelin-associated glycoprotein (MAG). These 'IgM anti-MAG positive MGUSP' patients characteristically present with slowly progressive, symmetric, predominantly distal sensory impairment with ataxia and little or no

weakness.[3] Since MGUS has an annual risk of 1% to transform to malignancy regular haematological screenings are recommended. Immuno-modulating treatment may be considered in patients with significant chronic or progressive disability, although none are proven effective. It is uncertain whether Rituximab has a beneficial effect in IgM anti-MAG positive MGUSP. Although several recent trials (including the yet unpublished full results of a randomised controlled trial from France) did not find improvement on the selected primary outcome measures after Rituximab in IgM-anti-MAG positive MGUSP, improvement was found on secondary outcome measures.[30, 31] This stresses the need to select a proper set of outcome measures for future clinical studies.

Multifocal motor neuropathy (MMN)

MMN is an infrequent, chronic disease and diagnosis is based on clinical, laboratory and electrophysiological characteristics.[2] Most patients with MMN have a slowly progressive or stepwise progression of asymmetric, predominantly distal weakness in the distribution of two or more peripheral nerves, for more than one month. Sensory symptoms are generally absent except for minor vibration sense abnormalities in the lower limbs. Other supportive features are predominantly upper limb involvement, decreased or absent tendon reflexes, absence of cranial involvement, the presence of cramps and fasciculations in the affected limb and response to immuno-modulatory treatment. The electrophysiological hallmark of MMN is persistent motor conduction block with reduction of the motor nerve conduction velocity only over the affected areas. Laboratory findings may reveal anti-ganglioside GM1 IgM antibodies in 30-80% of MMN patients. However these antibodies are not specific for MMN. Clinically, MMN is also described as an asymmetrical pure motor variant of CIDP with multifocal motor conduction blocks. Especially, during the evolution of MMN the multifocal character may gradually evolve in a more or less symmetrical pattern, clinically resembling the motor form of CIDP. Neuropathological studies have also linked MMN to CIDP.[32, 33] A multifocal acquired demyelinating sensory and motor neuropathy with conduction blocks (MADSAM or Lewis Sumner syndrome) has also been reported, hence fulfilling the intermediate clinical pattern between CIDP and MMN.[34, 35] IVIg is currently the standard treatment for patients with MMN.[2, 36]

Challenges in assessing outcome

In clinical trials involving patients with immune-mediated neuropathies outcome has been assessed using varying definitions of treatment response. Furthermore, many different outcome measures have been used, which may hamper comparison of trial results.[37] Therefore, the ***Peri*NomS** study was designed to improve and standardise assessment of patients with these conditions. The clinimetric essentials of selected outcome measures are being evaluated and compared to select a minimum core set of outcome measures for the future. This thesis also discusses the definition of treatment response by demonstrating the use of the concept of minimum clinically Important difference (MCID) in CIDP and the use of variable individual standard errors derived from Rasch-built interval scales in defining a responder.[38]

Table 1. Characteristics of typical GBS, CIDP, IgM anti-MAG MGUSP and MMN

	GBS	CIDP	IgM anti-MAG MGUSP	MMN
Clinical features	• Symmetric weakness of all extremities • Can be very severe (20-25% needs artificial ventilation) • Absent or reduced tendon reflexes • With or without sensory dysfunction • Cranial nerves may be affected • Sometimes autonomic dysfunction • Often pain	• Symmetric proximal and distal weakness of all extremities • Absent or reduced tendon reflexes • Often sensory dysfunction • Cranial nerves may be affected	• Symmetric, predominantly distal sensory impairment, often with ataxia • Generally little or no weakness • Tremor may be present	• Asymmetric, predominantly distal weakness in the distribution of two or more peripheral nerves (often starting in upper limb) • No sensory disturbances • No upper motor neuron signs • Absent or reduced tendon reflexes in affected limb • Absence of cranial nerve involvement
Course of disease	• Rapidly progressive onset • Maximum weakness within 4 weeks	• Deterioration continues > 2 months from onset • Often monophasic with stepwise progression • Sometimes relapsing with spontaneous remissions	• Slowly progressive onset • Chronic, duration over 6 months	• Slowly progressive or stepwise progressive onset • Chronic, duration usually over 6 months
Additional investigation	• EMG: demyelinating (e.g., AIDP) or axonal (e.g., AMAN) • CSF: increased protein level, normal cell count	• EMG: demyelinating, meeting specific electrophysiological criteria for CIDP • CSF: increased protein level, normal cell count	• EMG: demyelinating features similar to CIDP • Proven MGUS • IgM paraprotein: test for antibodies to MAG • CSF: sometimes increased protein level, normal cell count	• EMG: motor conduction blocks • Anti-GM1 IgM antibodies in 30-80% of MMN patients • CSF: protein level < 1 g/l, normal cell count
Treatment	• IVIg • Plasma exchange	• IVIg • Plasma exchange • Corticosteroids	• Immunosuppressive or immuno-modulating agents, (no proven effective drugs in RCTs)	• IVIg

Legend to table 1. GBS= Guillain-Barré syndrome, CIDP= chronic inflammatory demyelinating polyneuropathy, anti-MAG= antibodies against myelin-associated glycoprotein, MGUS= monoclonal gammopathy of undetermined significance, MGUSP= polyneuropathy associated with MGUS, MMN= multifocal motor neuropathy, EMG=electromyography, AIDP= acute inflammatory demyelinating polyradiculoneuropathy, AMAN= acute motor axonal neuropathy, CSF= cerebrospinal fluid, IVIg= intravenous immunoglobulins, RCTs= randomised controlled trials

References

1. Van den Bergh, P.Y., et al., *European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of chronic inflammatory demyelinating polyradiculoneuropathy: report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society - first revision.* Eur J Neurol, 2010. **17**(3): p. 356-63.
2. *European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of multifocal motor neuropathy. Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society--first revision.* J Peripher Nerv Syst, 2010. **15**(4): p. 295-301.
3. *European Federation of Neurological Societies/Peripheral Nerve Society Guideline on management of paraproteinemic demyelinating neuropathies. Report of a Joint Task Force of the European Federation of Neurological Societies and the Peripheral Nerve Society--first revision.* J Peripher Nerv Syst, 2010. **15**(3): p. 185-95.
4. van Doorn, P.A., L. Ruts, and B.C. Jacobs, *Clinical features, pathogenesis, and treatment of Guillain-Barré syndrome.* Lancet Neurol, 2008. **7**(10): p. 939-50.
5. CBO, *Multidisciplinaire richtlijn Guillain-Barré syndroom*. 2011.
6. Van der Meche, F.G., et al., *Diagnostic and classification criteria for the Guillain-Barré syndrome.* Eur Neurol, 2001. **45**(3): p. 133-9.
7. Asbury, A.K. and D.R. Cornblath, *Assessment of current diagnostic criteria for Guillain-Barré syndrome.* Ann Neurol, 1990. **27 Suppl**: p. S21-4.
8. Asbury, A., et al., *Criteria for diagnosis of Guillain-Barré syndrome.* Ann Neurol, 1978. **3**: p. 565-566.
9. Sejvar, J.J., et al., *Guillain-Barré syndrome and Fisher syndrome: case definitions and guidelines for collection, analysis, and presentation of immunization safety data.* Vaccine, 2011. **29**(3): p. 599-612.
10. Bernsen, R.A., et al., *Residual health status after Guillain-Barré syndrome.* J Neurol Neurosurg Psychiatry, 1997. **62**(6): p. 637-40.
11. Merkies, I.S., et al., *Connecting impairment, disability, and handicap in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2003. **74**(1): p. 99-104.
12. Meulstee, J. and F.G. van der Meche, *Electrodiagnostic criteria for polyneuropathy and demyelination: application in 135 patients with Guillain-Barré syndrome. Dutch Guillain-Barré Study Group.* J Neurol Neurosurg Psychiatry, 1995. **59**(5): p. 482-6.
13. Hughes, R., et al., *Subacute idiopathic demyelinating polyradiculoneuropathy.* Arch Neurol, 1992. **49**(6): p. 612-6.
14. Ruts, L., et al., *Distinguishing acute-onset CIDP from fluctuating Guillain-Barré syndrome: a prospective study.* Neurology, 2010. **74**(21): p. 1680-6.
15. Hughes, R.A., et al., *Immunotherapy for Guillain-Barré syndrome: a systematic review.* Brain, 2007. **130**(Pt 9): p. 2245-57.
16. Hughes, R.A., A.V. Swan, and P.A. van Doorn, *Intravenous immunoglobulin for Guillain-Barré syndrome.* Cochrane Database Syst Rev, 2010(6): p. CD002063.
17. Raphael, J.C., et al., *Plasma exchange for Guillain-Barré syndrome.* Cochrane Database Syst Rev, 2002(2): p. CD001798.
18. Gorson, K.C., G. Allam, and A.H. Ropper, *Chronic inflammatory demyelinating polyneuropathy: clinical features and response to treatment in 67 consecutive patients with and without a monoclonal gammopathy.* Neurology, 1997. **48**(2): p. 321-8.
19. Oh, S.J., J.L. Joy, and R. Kuruoglu, *"Chronic sensory demyelinating neuropathy": chronic inflammatory demyelinating polyneuropathy presenting as a pure sensory neuropathy.* J Neurol Neurosurg Psychiatry, 1992. **55**(8): p. 677-80.
20. *Research criteria for diagnosis of chronic inflammatory demyelinating polyneuropathy (CIDP). Report from an Ad Hoc Subcommittee of the American Academy of Neurology AIDS Task Force.* Neurology, 1991. **41**(5): p. 617-8.
21. *European Federation of Neurological Societies/Peripheral Nerve Society Guideline on management of chronic inflammatory demyelinating polyradiculoneuropathy: report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society--First Revision.* J Peripher Nerv Syst, 2010. **15**(1): p. 1-9.
22. Dyck, P.J., J. Prineas, and J. Pollard, *chronic inflammatory demyelinating polyneuropathy*, in *Peripheral neuropathy*. 1993, Saunders: Philadelphia. p. 1498-1518.

23. Dyck, P.J., et al., *Chronic inflammatory polyradiculoneuropathy.* Mayo Clin Proc, 1975. **50**(11): p. 621-37.
24. Dalakas, M.C., *Advances in the diagnosis, pathogenesis and treatment of CIDP.* Nat Rev Neurol, 2011.
25. Franssen, H., M. Vermeulen, and F.G. Jennekens, *Chronic inflammatory neuropathies*, in *Diagnostic criteria for neuromuscular disorders*, A.E.H. Emery, Editor. 1997, Royal society of medicine press: London. p. 53-59.
26. Eftimov, F., et al., *Intravenous immunoglobulin for chronic inflammatory demyelinating polyradiculoneuropathy.* Cochrane Database Syst Rev, 2009(1): p. CD001797.
27. Mehndiratta, M.M. and R.A.C. Hughes, *Corticosteroids for chronic inflammatory demyelinating polyradiculoneuropathy.* Cochrane Database Syst Rev, 2002(1): p. CD002062.
28. Mehndiratta, M.M., R.A.C. Hughes, and P. Agarwal, *Plasma exchange for chronic inflammatory demyelinating polyradiculoneuropathy.* Cochrane Database Syst Rev, 2004(3): p. CD003906.
29. RMC trial group, *Randomised controlled trial of methotrexate for chronic inflammatory demyelinating polyradiculoneuropathy (RMC trial): a pilot, multicentre study.* Lancet Neurol, 2009. **8**(2): p. 158-64.
30. Niermeijer, J.M., et al., *Rituximab for polyneuropathy with IgM monoclonal gammopathy.* J Neurol Neurosurg Psychiatry, 2009. **80**(9): p. 1036-9.
31. Leger, J.M., et al., *A randomized placebo-controlled trial of rituximab in IgM anti-myelin-associated glycoprotein antibody demyelinating neuropathy (RIMAG study).* J Peripher Nerv Syst, 2011. **16**(supplement 3): p. S73-S74.
32. Oh, S.J., et al., *Multifocal demyelinating motor neuropathy: pathologic evidence of 'inflammatory demyelinating polyradiculoneuropathy'.* Neurology, 1995. **45**(10): p. 1828-32.
33. Krendel, D.A., *Biopsy findings link multifocal motor neuropathy to chronic inflammatory demyelinating polyneuropathy.* Ann Neurol, 1996. **40**(6): p. 948-50.
34. Sharoqi, I.A., et al., *Chronic relapsing multifocal sensory-motor neuropathy with conduction block.* J Peripher Nerv Syst, 1998. **3**(2): p. 133-6.
35. Lewis, R.A., et al., *Multifocal demyelinating neuropathy with persistent conduction block.* Neurology, 1982. **32**(9): p. 958-64.
36. Eftimov, F. and I.N. Van Schaik, *Immunotherapy of multifocal motor neuropathy.* Expert Opin Biol Ther, 2011. **11**(3): p. 329-42.
37. van Nes, S.I., C.G. Faber, and I.S. Merkies, *Outcome measures in immune-mediated neuropathies: the need to standardize their use and to understand the clinimetric essentials.* J Peripher Nerv Syst, 2008. **13**(2): p. 136-47.
38. Merkies, I.S., et al., *Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance.* J Neurol Neurosurg Psychiatry, 2010. **81**(11): p. 1194-9.

Chapter 1.2

Outcome measures in immune-mediated neuropathies: the need to standardise their use and to understand the clinimetric essentials

Review

Sonja I. van Nes[1], Catharina G. Faber[2], and Ingemar S.J. Merkies[1,2,3]

[1]Department of Neurology, Erasmus Medical Centre Rotterdam, the Netherlands
[2]Department of Neurology, Maastricht University Medical Centre, Maastricht, the Netherlands
[3]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands

J Peripher Nerv Syst. 2008 Jun;13(2):136-47

Abstract

Peripheral neurological disorders like neuropathies may cause impairments (such as weakness and sensory deficits) which may lead to problems in daily life and social functioning with a possible decrement in quality of life expectations. Choosing the proper outcome measure to evaluate the therapeutic efficacy of an intervention at one of these levels of outcome should therefore be considered as fundamental to the design of randomised trials in peripheral neurological disorders. However, these choices are not only dependent of the proposed research purposes, but also, and perhaps more importantly, of the fulfilment of the scientific needs of these measures. With an increasing demand for accuracy, a thorough and comprehensive evaluation of an outcome measure is needed to determine its simplicity, communicability, validity, reliability, and responsiveness before being clinically applicable; techniques which are being captured by the science of clinimetrics. Most neurologists are still unfamiliar with these rigorous methodological essentials or overlook some of them in their trial preparations because these are considered time consuming and mind numbing.

This review will highlight, against the background of the international classification framework and clinimetric needs for outcome measures, the selected scales applied in published randomised controlled trials in patients with Guillain-Barré syndrome, chronic inflammatory demyelinating polyradiculoneuropathy, multifocal motor neuropathy, and gammopathy related neuropathies. The need for comparison responsiveness studies between equally valid and reliable measures and to standardise their use is emphasised in these conditions. Finally, specific recommendations are given to move from classic to modern clinimetric approaches when constructing, evaluating, and selecting outcome measures using new methods like Rasch analysis, accentuating the need for shifting towards a more modern era.

Introduction

The American Academy of Neurology, the American Association of Electrodiagnostic Medicine, and the American Academy of Physical Medicine and Rehabilitation recently stated that the combination of neuropathic symptoms, signs, and electrodiagnostic findings should provide the most accurate diagnosis of distal symmetric peripheral neuropathies (PN). However, the recommendations hardly addressed the (lack of) standardisation of assessment of symptoms and signs.[1] Moreover, the translation of symptoms and signs in these conditions to limitations in daily activities and social participation with possible decrement in quality of life expectations was not addressed. Neither was the need for standardisation of assessing outcome at these various levels referred. It is therefore not surprising that an overwhelming assortment of scales have been applied in clinical therapeutic trials in PN.[2] Some scales were introduced before they had been fully clinimetrically tested. Others consisted of a mixture of different clinical modalities.[3-5] Flawed measures threaten the significance of trials that use them and impede comparison of results.

With an increasing demand for accuracy, selected outcome measures need to be clinimetrically well evaluated meeting the demands of being simple, communicable, valid, reliable, and responsive.[3-5] Moreover, outcome measures should be unambiguously constructed to represent only one of the outcome levels according to the international classification by the World Health Organisation (WHO) and quality of life concept.[6-8] These aspects are considered cardinal features in the evaluation of outcome measures.

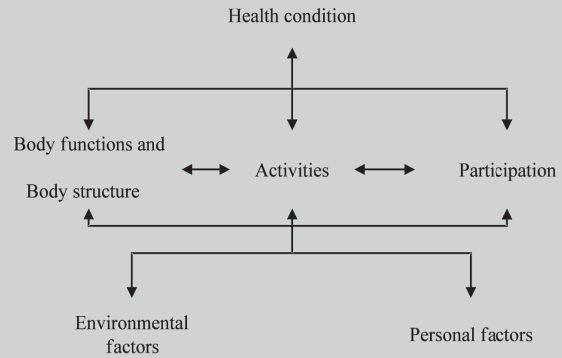
This paper provides an overview of all selected outcome measures applied in clinical randomised trials of patients with an immune-mediated neuropathies like Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), multifocal motor neuropathy (MMN), and monoclonal gammopathy of undetermined significance related neuropathy (MGUSP). The significance of these selected outcome measures will be reflected against the background of their levels of representation in accordance with the WHO framework and their fulfilment of clinimetric essentials. Finally, recommendations will be given, emphasising the need to standardise their use and to adopt the modern rather than the classical clinimetric approach when constructing outcome measures for peripheral nerve diseases.

International classification of outcome measures – the WHO principles and quality of life concept

The 'consequences' of chronic illnesses have gained significant attention due to their life-long management needs. In particular, the management of these conditions became the goal and the use of outcome measures became the standard for measuring the performance of health care delivery and its effectiveness. Therefore, in 1980 the WHO published a framework, the International Classification of Impairments, Disabilities, and Handicaps (ICIDH), to structure outcome measures used in evaluating the consequences of illnesses.[7] This international classification was renewed in 2001 as the International Classification of Functioning, Disability, and Health (ICF), providing more information on the relationship between the various levels of outcome from a more multidimensional point of view taking the possible influence of environmental and personal factors on health conditions into account.[8] Figure 1 provides the structure of the ICF. The ICF is organised in two parts. The first part recognises two main components of functioning and disability, namely: a) an impairment component comprising 'body function' and 'body structures', and b) an activities and participation component providing a complete set of domains for aspects of daily and social functioning. In the second part, the impact of environmental and personal factors is presented, thus showing a dynamic interaction between health and environmental and personal factors. The ICF serves as a conceptual framework to bring together the physiological, personal and societal aspects of consequences related to health conditions.[8]

Another aspect to take into consideration is the assessment of outcome from the 'patient's own perspective', a concept captured in 'quality of life' assessment.[6] Quality of life is defined as the patient's reaction to the discrepancy between actual and expected achievements arising as a consequence of illness. At least four dimensions should be included in a quality of life assessment. These dimensions are physical, functional, psychological, and social health. The physical health dimension refers primarily to disease- and treatment-related symptoms. Functional health comprises self-care, mobility, and physical activity level, as well as the capacity to carry out various roles in relation to family and work. Cognitive functioning, emotional status, and general perceptions of health, well-being, life satisfaction, and happiness are the central components of the psychological health domain. Social functioning includes the assessment of qualitative and quantitative aspects of social contacts and interactions.[6] Hence, in the preparation of any randomised study, it is essential to first determine which level will be of primary interest to answer the research question(s) and secondly to choose an outcome measure that represents that particular level of outcome and which has demonstrated good clinimetric soundness.

Figure 1. International Classification of Functioning, Disability and Health (ICF)



Legend to figure 1. The following definitions are provided for the various components of the ICF: *Body functions*: physiological and psychological functions of body systems; an example is weakness or fatigue; *Body structures*: anatomical parts of the body such as organs, limbs and their components; examples are structures of the peripheral nervous system; *Impairments* are considered problems in body function or structure as a significant deviation or loss; *Activity*: the execution of a task or action by an individual; *Participation*: the involvement of an individual in a life situation. *Environmental factors* are extrinsic to (outside of) the individual (e.g., the attitude of society, architectural characteristics, the legal system). *Personal factors* are intrinsic to (inside of) the individual and describe on how the consequences of a health status is experienced (e.g., gender, age, fitness, lifestyle, habits, coping styles, social background, education, life experience).[8]

Scientific evaluation of outcome measures in immune-mediated neuropathies: the clinimetric essentials

For proper measurement in clinical practice, outcome measures should fulfil all 'clinimetric' properties prior to their use. The term 'clinimetrics' was introduced by Alvan R. Feinstein in 1987 to indicate a set of methods, strategies, and principles for evaluating and constructing outcome measures.[3] He stated in his introduction that '... patients and clinicians ... may not realize how often they communicate with *clinimetric indexes* ... like *severe* pain, a *slight* fatigue, a *great improvement* ...', and that '... these expressions are seldom regarded as acts of measurement'. Investigation of outcome measures involves the evaluation of their clinical applicability and scientific soundness. These entities are captured in the various clinimetric properties like simplicity, communicability, validity, reliability, and responsiveness.[3-5] The strength of each of these properties should be extensively determined.

A clinical useful scale should fulfil the following requirements:

Simplicity
An outcome measure should be *simple, none-time consuming with little special training*. Many outcome measures are impractical because they require too much time to administer or score. A measure should wherever possible be simple, particularly if more than one person is going to use the measure. Simplicity will improve the patient (and user) compliance, and will increase reliability.

Communicability
An outcome measure should provide *results that easily can be interpreted by others (reflecting good communicability)*. A measure should give results that are easily and unambiguously understood by others. The construction of outcome measures should also contain unambiguously interpretable items.

Validity
An outcome measure should be *valid*. A valid scale is one that *measures what it purports to measure* and therefore provides the information required. In other words: It should accurately describe the underlying phenomenon or disease. Various types of validity are described:
Face validity refers to the apparent sensibility of the measure and its components. It simply indicates whether, on the face of it, the scale appears to be assessing the desired qualities. This validity form represents the subjective judgment based on a review of the measure itself by one or more experts, and rarely are any empirical approaches used. The GBS disability score has obvious face validity for the assessment of mobility.[9]
Content validity is closely related to the face validity concept. It consists of a judgment by experts evaluating whether an outcome measure captures all the relevant or important contents or domains of an illness. The GBS disability scale (also known as Hughes' functional grading scale or f-score) is a 7-point disability scale ranging from no symptoms (zero points) to death (six points).[9] This scale would be a content valid measure for lower limb function in neuropathies, because it is strongly based upon mobility. These two forms of validity are also entitled as 'the validity forms by assumption', meaning that a measure assesses outcome in a certain way because an expert says it does.[3-5]
Construct validity is demonstrated by examining the relations between a newly created test and other tests to show that the new test measures the same 'construct'. In practice, evidence for construct validity is gathered by undertaking a series of studies to determine:
- *Convergent validity* – the extent to which a measure correlates with other measures of related entities.
- *Discriminant validity* – the extent to which a measure does not correlate with measures of different entities.
- *Divergent validity* – the extent to which a measure correlates with measures of opposite entities (e.g., correlation between fatigue versus vitality scales).

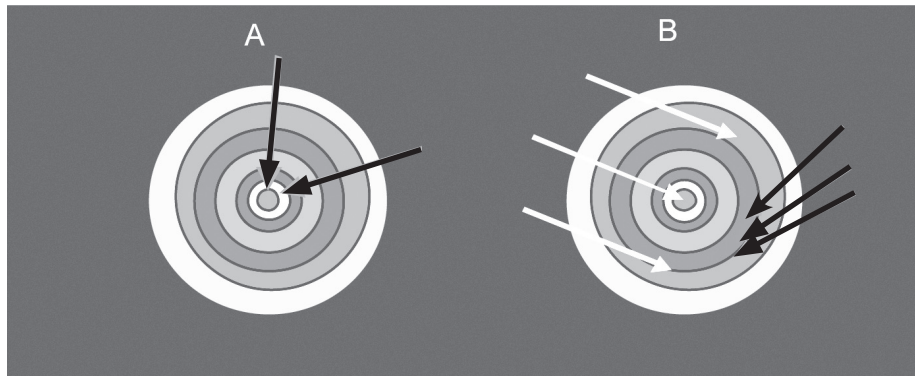
Criterion-related validity is demonstrated by examining the accuracy of a test compared with a particular standard, the criterion ('gold standard'). There are two types of criterion-related validity:
- *Concurrent validity* – the extent to which a new measure correlates with another widely accepted validated measure or the opinion of experts. This is generally applied in cases of no 'real' gold standard.
- *Predictive validity* – if we thought that the GBS disability score values at 4 weeks of follow-up in patients with GBS could predict degree of disability at 6 months, data collected at these two instances in time must be correlated.

Reliability
An outcome measure should be *reliable*. A reliable measure is one that produces *results that are accurate, consistent, stable over time, and reproducible*. A patient whose condition has not changed should always receive the same score apart from random variation. There are three different types of reliability:
Internal consistency (Interitem consistency) is the extent to which items comprising a scale measure the same concept – that is, a measure of the homogeneity of the scale. There are a number of ways to calculate these correlations, of which the Cronbach's alpha is the most widely used.[10] After relatedness of items has been excluded with factor analysis, an alpha ≥ 0.7 is considered having a good internal consistency, but this value alone will not suffice in determining the reliability of an outcome measure.
Observer reliability is the agreement between observers or within an individual observer. There are two types:
- *Interobserver reliability* – the agreement between observations made by two or more raters on the same patient or group of patients.
- *Intraobserver reliability* – the agreement between observations made by the same rater on two different occasions on the same patient or group of patients.

Test-retest reliability is the agreement between observations made by the same patient on two different occasions.

The concepts of *validity* and *reliability* can be explained using the example of 'shooting at a target'. Someone learning archery must first learn to hit the centre of the target, and then to do this consistently (figure 2).

Figure 2. The concept of validity and reliability explained by the example of shooting on a target



Legend to figure 2. Validity is represented by the aim of shooting; in other words, how close the shots come to the centre of the target. Good validity meaning a bias of approximately zero distance (panel A). Reliability is represented by how close consecutive shots fall to each other wherever they land. Good reliability means a small variance. The white arrows in panel B are more scattered (= lower reliability) compared with the black arrows. However, the validity of the black arrows in panel B is low, since they do not hit the aiming spot (centre of the target).

Responsiveness

An outcome measure should be *responsive to changes over time in the underlying condition, yet relatively insensitive to minor symptom and sign fluctuations.* Whereas validity and reliability form the clinimetric foundation of a rating scale, the ability of a measure to detect clinically meaningful changes over time is crucial. For clinicians and researchers, such a measure should discriminate between irrelevant changes (normal fluctuations in the activity of an illness; 'noise') and clinically meaningful changes on which a treatment policy can be based ('signal'), an ability addressed as 'responsiveness'. A statistic and heuristic approach in examining responsiveness of a measure has been proposed.[11]

- Statistical responsiveness captures the ability of an instrument to measure any change, irrespective of its relevance. Techniques that capture statistical responsiveness are the effect size, standardised response mean, Guyatt's responsiveness ratio in cases of parametric statistics, and the Schmitz' distribution-free responsiveness score for non-parametric data.[12-14]
- Heuristic techniques are based upon comparing changes as assessed by a scale with an external indicator, for example the grades of judgment by the patients of their clinical condition compared to the last consultation visit (e.g., grade 1: improved; grade 2: stable; grade 3: deteriorated).

Overview of applied outcome measures in randomised clinical trials in patients with immune-mediated neuropathies

Methods

A systematic search was performed using the search engines of Medline, EMBASE, Cochrane Neuromuscular Disease Group (period: January 1976 to March 2008) to detect all randomised clinical studies in patients with GBS, CIDP, MMN, and MGUSP. Selection was based on the following keywords: Guillain-Barré syndrome, GBS, chronic inflammatory demyelinating poly(radiculo)neuropathy, CIDP, inflammatory (poly)(radiculo)neuropathy, immune-mediated polyneuritis, gammopathy, dysimmune, paraprotein(a)emia, and monoclonal gammopathy of undetermined significance polyneuropathy, controlled clinical trial and randomised controlled trials. Studies included in this review were selected independently by two authors (SvN and IM). All measures, primary and secondary, used to evaluate the effect of therapeutic interventions were categorised according to the renewed WHO international classification and are presented in the tables 1A – 1D.[8, 15] Various quality-control meetings were performed to improve categorisation.

Table 1A. Outcome measures used in randomised trials in patients with GBS

	year	authors
Impairment measures		
Muscle strength measures		
Muscle strength	1984	Osterman et al.[16]
	1985	Mendell et al.[17]
Hand grip strength	1987	Farkkila et al.[18]
MRC sumscore	1992	Van der Meché et al.[19]
	2004	Van Koningsveld et al.[20]
Strength recovery	1995	Gürses et al.[21]
Time to onset of motor recovery	1997	French cooperative group on PE in GBS[22]
Sensory measures		
Sensory scale	1992	Van der Meché et al.[19]
Sensory symptoms	2004	Van Koningsveld et al.[20]
Other measures		
Time to recover	1976	Swick and McQuillen[23]
Forced vital capacity	1985	Mendell et al.[17]
Functional test	1985	Mendell et al.[17]
Reduction of time to cease artificial ventilation	1993	GBS steroid trial group[24]
Time to discontinuation of ventilation	1997	PSGBS Trial group[25]
Numeric pain rating scale	2002	Pandey et al.[26]
Ramsey sedation score	2002	Pandey et al.[26]
Cranial nerve (dys)function	2004	Van Koningsveld et al.[20]
	2005	Korinthenberg et al.[27]
Respiratory function	2005	Korinthenberg et al.[27]
Vegetative symptoms	2005	Korinthenberg et al.[27]
Pain	2005	Korinthenberg et al.[27]
Arm function	2005	Korinthenberg et al.[27]
Fatigue severity scale	2006	Garssen et al.[28]
Fatigue impact scale	2006	Garssen et al.[28]
Hospital anxiety and depression scale	2006	Garssen et al.[28]
Activity and participation measures		
GBS disability score	1978	Hughes et al.[9]
	1984	Greenwood et al.[29]
	1985	GBS study group[30]
	1988	Shukla et al.[31]
	1992	Van der Meché et al.[19]
	1993	GBS steroid trial group[24]
	1996	Bril et al.[32]
	1996	Haupt et al.[33]

Table 1A. (Continued)

	year	authors
Activity and participation measures		
GBS disability score	1996	Singh and Gupta[34]
	1997	PSGBS Trial group[25]
	2001	Wollinsky et al.[35]
	2000	Nomura et al.[36]
	2001	Wang et al.[37]
	2003	Pritchard et al.[38]
	2004	Van Koningsveld et al.[20]
	2005	Korinthenberg et al.[27]
	2007	Garssen et al.[39]
Time to recover ability to do manual work	1978	Hughes et al.[9]
	1996	Bril et al.[32]
Nine point disability grade	1984	Osterman et al.[16]
Time to recover walking with assistance	1987 + 1997	French cooperative group on PE in GBS[22, 40]
	2001	Raphael et al.[41]
Time to recover unaided walking	1995	Gürses et al.[21]
Time to walk unaided	1997	PSGBS Trial group[25]
Rotterdam handicap scale	2006	Garssen et al.[28]
Quality of life measures		
SF-36	2006	Garssen et al.[28]
EuroQol-5D	2006	Garssen et al.[28]

Table 1B. Outcome measures used in randomised trials in patients with CIDP

	year	authors
Impairment measures		
Muscle strength measures		
MRC sumscore	1993	Vermeulen et al.[42]
	2001	Hughes et al.[43]
	2008	Hughes et al.[44]
Jamar dynamometer	1996	Hahn et al.[45]
	1996	Hahn et al.[46]
Expanded MRC sumscore	1999	Hadden et al.[47]
Average muscle score	2001	Mendell et al.[48]
Vigorimeter	2001	Hughes et al.[43]
	2008	Hughes et al.[44]
Sensory measures		
INCAT sensory sumscore	2001	Hughes et al.[43]
	2008	Hughes et al.[44]
Composite and other measures		
NIS	1985	Dyck et al.[49]
	1996	Hahn et al.[46]
	1996	Hahn et al.[45]
Forced vital capacity	2001	Mendell et al.[48]
Activity and participation measures		
Rankin scale	1990	Van Doorn et al.[50]
	1993	Vermeulen et al.[42]
	2001	Hughes et al.[43]
Functional clinical grading scale	1996	Hahn et al.[45]
	1996	Hahn et al.[46]
Ten meters walking test	1999	Hadden et al.[47]
	2001	Hughes et al.[43]
Nine hole pegboard test	1999	Hadden et al.[47]
	2001	Hughes et al.[43]
Ambulation index	1999	Hadden et al.[47]
Functional independence measure	1999	Hadden et al.[47]
Hammersmith motor ability	1999	Hadden et al.[47]
Guy's neurological disability scale	1999	Hadden et al.[47]
GBS disability score	2001	Mendell et al.[48]
Rotterdam Handicap scale	2001	Hughes et al.[43]
INCAT Overall disability sumscore	2001	Hughes et al.[43]
	2008	Hughes et al.[44]
Quality of life measures		
EuroQol-5D	1999	Hadden et al.[47]
SF-36	2001	Hughes et al.[43]

Legend to tables 1A-1D. MRC=Medical Research Council, INCAT=Inflammatory Neuropathy Cause and Treatment, NIS=neuropathy impairment score, SF-36=short form 36-item health survey.

Table 1C. Outcome measures used in randomised trials in patients with MMN

	year	authors
Impairment measures		
Muscle strength measures		
Manual muscle strength testing	1995	Van den Berg et al.[51]
Hand-held dynamometry	2007	Piepers et al.[52]
Jamar dynamometer	1995	Van den Berg et al.[51]
Subjective rating of strength	2007	Piepers et al.[52]
MRC score	2000	Federico et al.[53]
Grip strength	2000	Federico et al.[53]
	2001	Léger et al.[54]
	2007	Piepers et al.[52]
Composite measures		
NIS	2000	Federico et al.[53]
Activity and participation measures		
Rankin scale	1995	Van den Berg et al.[51]
Subjective rating of functionality	2000	Federico et al.[53]
Self-evaluation scale	2001	Léger et al.[54]
	2007	Piepers et al.[52]
Nine hole pegboard test	2007	Piepers et al.[52]
Guy's disability arm grade	2007	Piepers et al.[52]

Table 1D. Outcome measures used in randomised trials in patients with MGUSP

	year	authors
Impairment measures		
Muscle strength measures		
Manual muscle strength testing	1996	Dalakas et al.[55]
	2007	Niermeijer et al.[56]
MRC sumscore	2001	Comi et al.[57]
Sensory measures		
Sensory score	1996	Dalakas et al.[55]
Subjective assessment of sensation	1997	Mariette et al.[58]
INCAT sensory sumscore	2001	Comi et al.[57]
Vibration threshold	2001	Comi et al.[57]
Sensory scale testing	2007	Niermeijer et al.[56]
Ataxia tapping test	2007	Niermeijer et al.[56]
Composite measures		
NIS	1991	Dyck et al.[59]
	1995	Oksenhendler et al.[60]
	1997	Mariette et al.[58]
Activity and participation measures		
Neuromuscular symptom score	1996	Dalakas et al.[55]
INCAT overall disability sumscore	2001	Comi et al.[57]
Ten meters walking test	2001	Comi et al.[57]
Nine hole pegboard test	2001	Comi et al.[57]
Rotterdam handicap scale	2001	Comi et al.[57]
Rivermead mobility index	2007	Niermeijer et al.[56]
Rankin scale	2007	Niermeijer et al.[56]
Quality of life measures		
SF-36	2001	Comi et al.[57]
	2007	Niermeijer et al.[56]

Results

There was a 100% match on selected randomised clinical studies in these conditions by the two authors that performed the search. In patients with GBS, most impairment scales were based on the MRC grading.[61] Of these, only the MRC sumscore has demonstrated the complete clinimetric arsenal of being simple, valid, reliable, and responsive.[62] Some impairment measures were based on neurological examination findings. Of the fatigue scales used, the fatigue severity scale has demonstrated its simplicity, communicability, validity, and reliability in patients with immune-mediated neuropathies.[63] To date, no responsiveness studies using this scale have been conducted. At the activity and participation level, most outcome measures captured mobility aspects of which the GBS disability score have been most widely used. This scale has a strong face and content validity for mobility aspects, but fails in assessing arm functionality. All clinimetric properties of this scale have been demonstrated. Surprisingly, only three randomised studies in GBS have addressed arm function, but all failed in providing the scientific soundness of these outcome measures before their use (see table 1A).[9, 27, 32]

In patients with CIDP, the neuropathy impairment score (NIS; formerly named neuropathy disability score) has been most widely used.[64] The validity, reliability, and responsiveness of this composite score have been demonstrated. However, the NIS has a poor communicability, since it is composed by various neurological items like different sensory modalities, strength assessment, and tendon reflex findings. The complexity of such a score makes interpretation of changes in the final score difficult. Moreover, the significance of possible changes in tendon reflex scores has been questioned, since these changes are hardly translated into functional deficit.

At the activity and participation level, a variety of outcome measures have been applied, most of these being strongly directed towards mobility. Surprisingly, only the Inflammatory Neuropathy Cause and Treatment (INCAT) overall disability sumscore (ODSS) has provided a 'true general outcome' (defined as functional arms + legs information) in patients with immune-mediated polyneuropathies and has recently demonstrated its efficacy in the largest trial ever conducted in patients with CIDP being treated with intravenous immunoglobulin (see table 1B).[44, 65]
Although not listed, the overall neuropathy limitations scale (ONLS) is currently being applied in an ongoing CIDP trial. The ONLS also addresses arms and legs functionality and has demonstrated good scientific soundness.[66]

In MMN and MGUSP the same pattern of great variety of outcome measures is listed (tables 1C and 1D). Moreover, some outcome measures have been introduced without being first clinimetrically evaluated. An example of the latter is the recently reported randomised trial by Niermeijer et al. (2007) who examined the efficacy of intermittent cyclophosphamide with prednisone in patients with MGUSP.[56] These researchers chose the Rivermead mobility index (RMI) to capture possible efficacy

as the primary endpoint, but have failed to examine whether this outcome measure would have been communicable, valid, reliable, and last but not least responsive to capture relevant changes over time in this condition. As a result, while impairment measures with good clinimetric soundness demonstrated relevant improvement over time, the RMI did not.[56] A more appropriate activity and participation scale might have captured changes at this level, thus changing the final results of this trial.

Surprisingly, only a limited number of trials in patients with immune-mediated neuropathies have addressed the concept of quality of life. The EuroQol-5D and the SF-36 have been applied in these studies.[28, 43, 47, 56, 57] Both scales have demonstrated their scientific soundness in these conditions, but the SF-36 has been examined more extensively and has shown its complementary qualities to traditional outcome measures.[67]

Positioning outcome measures in immune-mediated neuropathies: where do we stand now and what are the future needs?

Recognising the need to standardise the use of outcome measures

As the tables illustrate, an overwhelming collection of outcome measures has been applied in these conditions, which hamper comparison between the obtained results. Surprisingly, only recently attention was directed towards standardisation of the use of outcome measures at all levels of outcome in inflammatory polyneuropathies.[2, 68] At the 2004 meeting in the Netherlands, *general* consensus regarding the use of various outcome measures was reached. However, further recommendations were made to enhance the clinimetric comparability between the various selected outcome measures and some of these are briefly highlighted:

- In diseases such as in MMN, outcome measures need to be evaluated or developed covering each of the ICF domains.
- In diseases in which symptoms are prominent, such as painful neuropathy, the use of symptom-based outcome measures should be considered.
- Evaluators should be trained in the application of the selected instruments before the start of each study.
- Research should be conducted to compare the responsiveness of valid and reliable outcome measures recommended in different patient populations. In GBS, CIDP, and gammopathy related neuropathies, comparison is needed between MRC sumscore versus NIS motor subset and between INCAT sensory sumscore versus NIS sensory subset.[62, 64, 69]
- Assessment of both quality of life with SF-36 and health utility with the EuroQol should be considered.[70, 71]

Shifting from classic to modern clinimetric approach

By consensus among the INCAT group, the activity and participation level was proposed as the main level of measuring therapeutic response and the ODSS was proposed as a preferential outcome measure for treatment trials of these conditions, since it covers not only mobility disturbances, but also upper limb dysfunction[2]. Also, the ODSS has good clinimetric properties, especially greater responsiveness and has demonstrated a higher proportion of variance of disability explained by impairment qualities.[72] In addition to these findings, higher associations were found between ODSS and outcome assessed from patients' perceptions in immune-related polyneuropathies than in other commonly used disability scales.[73]

Despite these findings, all disability measures (including the ODSS) used in immune-mediated polyneuropathies (see tables 1A-1D), are criticised based on the following:

- The construction of these ordinal (or Likert) outcome measures is based on the *classic* test theory which recruits items arbitrarily without weighting or standardisation.
- Patients are requested to complete all the items in a classic outcome measure, even though some may be inappropriate for their level of ability. Hence, it was suggested that classically derived tests are only suitable for group studies but not suitable for individual patient assessments.
- Under classic test theory, the patient's raw test score would be the sum of the scores received on the items in the test. The sumscore assumes that all items have equal relevance and weights regarding the functionality of patients and that given differences in sumscore have equal meaning. Therefore, classic test theory hampers the comparison between patients with different degrees of functionality.
- Hundreds of multi-item disability scales have been developed for various chronic illnesses in the last decades. Unfortunately in defining these scales, researchers have not attempted to define these outcome measures from the patient's perspective.
- It is questioned whether so many scales are necessary, since activity and participation limitations are considered relatively independent of underlying illnesses.
- A large pool of disability scales hampers comparison of the results of clinical trials, since the clinical importance of score changes on all scales is unknown.[74, 75]

It is clear that a new calibrated linearly weighted interval disability measure is needed using modern scientific methods. The Rasch technology and item response theory (IRT) are increasingly being recognised and applied as modern test methods.[76, 77] These statistical techniques attempt to transform ordinal scores, that are scale dependent and of limited accuracy, into interval measures that are scale independent

and suitably accurate for individual patient assessment. In essence, these methods model the probability of an individual's response to an item. They are based on a logical assumption: individuals with high ability to perform a task should have an increased probability, relative to individuals with low levels, of getting a better score on any item.[76, 77] Using these new techniques would give a true reflection of disease impact, differences between individuals and groups, and treatment effects. The ability to generate interval measures, independent of the rating scale used, enables scales measuring the same health construct to be equated on the same linear ruler. This is the basis for comparisons of studies, systematic reviews, and meta-analyses. Moreover, the process of scale equating generates a pool of commonly calibrated items, which form an item bank. Item banks are flexible measurement methods because any subset of items can be selected from the bank to generate an accurate score.[78] With these techniques, investigators are no longer wedded to defined scales and can simply select the most appropriate group of items for their study. Alternatively, a fixed defined group of items can be selected for general use in a particular illness. This would be, for example, of great interest when choosing items that can be used in a clinical trial evaluating improvement in patients with mild forms of GBS.

The use of Rasch technology is suggested rather than traditional methods and IRT for very specific reasons: the aim of a Rasch analysis is to determine the extent to which the observed rating scale data satisfy the requirements of the mathematical measurement model. When the observed data satisfy these requirements, within reason, scores generated by ordinal rating scales can be transformed into interval level measurements. In contrast, IRT models are derived to explain data.[79, 80] It is now up to the neurology community to learn the Rasch theory and practice of it, and to present it in an understandable way.

New international clinimetric study in immune-mediated neuropathies

Based on the above, it is concluded that consensus regarding the use of a *specific* core set of outcome measures is urgently needed to improve the assessment of new drug studies in these PN forms. Therefore, the international multi-centre *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study has been conceived as part of the Inflammatory Neuropathy Consortium (INC) aiming to evaluate equally valid and reliable outcome measures of interest through comparative responsiveness. The construction of a linearly weighted interval activity and participation multi-item scale based on Rasch methodology will also be part of the ***P*eri*N*om*S*** study.

Conclusion

This review has shown the diversity of selected outcome measures in published randomised controlled trials in patients with immune-mediated neuropathies. Choosing proper outcome measures to evaluate the therapeutic efficacy of an intervention should be considered for the design of randomised trials in peripheral neurological disorders. The selected outcome measures need to fulfil the clinimetric essentials like being simple, communicable, valid, reliable, and responsive. Because ordinally driven outcome measures have serious weaknesses, we propose to move from classic to modern clinimetric approach when constructing, evaluating, and selecting outcome measures. The Rasch analysis method is preferred for this purpose. Finally, there is a strong plea for standardisation and the use of a core set of outcome measures at every level of outcome in all randomised trials in immune-mediated neuropathies, promoting comparability of the obtained results.

References

1. England, J.D., et al., *Distal symmetrical polyneuropathy: a definition for clinical research. A report of the American Academy of Neurology, the American Association of Electrodiagnostic Medicine, and the American Academy of Physical Medicine and Rehabilitation.* Arch Phys Med Rehabil, 2005. **86**(1): p. 167-74.
2. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
3. Feinstein, A.R., *Clinimetrics*. 1987, New Haven and London: Yale University Press.
4. Hobart, J.C., D.L. Lamping, and A.J. Thompson, *Evaluating neurological outcome measures: the bare essentials.* J Neurol Neurosurg Psychiatry, 1996. **60**(2): p. 127-30.
5. Streiner, D.L.,and G.R. *Norman Health Measurement Scales. A Practical Guide to their development and use.* 1998, 2nd Edn. New York: Oxford University Press
6. Aaronson, N.K., *Quality of life: what is it? How should it be measured?* Oncology (Williston Park), 1988. **2**(5): p. 69-76, 64.
7. World Health Organisation, *International classification of impairments, disabilities, and handicaps*. 1980: Geneva.
8. World Health Organisation, *International classification of impairments, disabilities, and handicaps*. 2001: Geneva.
9. Hughes, R.A., et al., *Controlled trial prednisolone in acute polyneuropathy.* Lancet, 1978. **2**(8093): p. 750-3.
10. Nunnally, J.C., *Psychometric theory*. 1978, New York: McGraw Hill.
11. Liang, M.H., *Evaluating measurement responsiveness.* J Rheumatol, 1995. **22**(6): p. 1191-2.
12. Kazis, L.E., J.J. Anderson, and R.F. Meenan, *Effect sizes for interpreting changes in health status.* Med Care, 1989. **27**(3 Suppl): p. S178-89.
13. Liang, M.H., A.H. Fossel, and M.G. Larson, *Comparisons of five health status instruments for orthopedic evaluation.* Med Care, 1990. **28**(7): p. 632-42.
14. Merkies, I.S., et al., *Assessing grip strength in healthy individuals and patients with immune-mediated polyneuropathies.* Muscle Nerve, 2000. **23**(9): p. 1393-401.
15. Fitzpatrick, R., et al., *Evaluating patient-based outcome measures for use in clinical trials.* Health Technol Assess, 1998. **2**(14): p. i-iv, 1-74.
16. Osterman, P.O., et al., *Beneficial effects of plasma exchange in acute inflammatory polyradiculoneuropathy.* Lancet, 1984. **2**(8415): p. 1296-9.
17. Mendell, J.R., et al., *Plasma exchange and prednisone in Guillain-Barré syndrome: a controlled randomised trial.* Neurology, 1985. **35**(11): p. 1551-5.
18. Farkkila, M., et al., *Guillain-Barré syndrome: quantitative measurement of plasma exchange therapy.* Neurology, 1987. **37**(5): p. 837-40.
19. van der Meche, F.G. and P.I. Schmitz, *A randomised trial comparing intravenous immune globulin and plasma exchange in Guillain-Barré syndrome. Dutch Guillain-Barré Study Group.* N Engl J Med, 1992. **326**(17): p. 1123-9.
20. van Koningsveld, R., et al., *Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barré syndrome: randomised trial.* Lancet, 2004. **363**(9404): p. 192-6.
21. Gurses, N., et al., *Intravenous immunoglobulin treatment in children with Guillain-Barré syndrome.* Scand J Infect Dis, 1995. **27**(3): p. 241-3.
22. French Cooperative Group on Plasma Exchange in Guillain-Barré Syndrome, *Appropriate number of plasma exchanges in Guillain-Barré syndrome.* Ann Neurol, 1997. **41**(3): p. 298-306.
23. Swick, H.M. and M.P. McQuillen, *The use of steroids in the treatment of idiopathic polyneuritis.* Neurology, 1976. **26**(3): p. 205-12.
24. Guillain-Barré Syndrome Steroid Trial Group, *Double-blind trial of intravenous methylprednisolone in Guillain-Barré syndrome.* Lancet, 1993. **341**(8845): p. 586-90.
25. Plasma Exchange/Sandoglobulin Guillain-Barré Syndrome Trial Group, *Randomised trial of plasma exchange, intravenous immunoglobulin, and combined treatments in Guillain-Barré syndrome.* Lancet, 1997. **349**(9047): p. 225-30.

26. Pandey, C.K., et al., *Gabapentin for the treatment of pain in guillain-barre syndrome: a double-blinded, placebo-controlled, crossover study.* Anesth Analg, 2002. **95**(6): p. 1719-23, table of contents.
27. Korinthenberg, R., et al., *Intravenously administered immunoglobulin in the treatment of childhood Guillain-Barré syndrome: a randomised trial.* Pediatrics, 2005. **116**(1): p. 8-14.
28. Garssen, M.P., et al., *Amantadine for treatment of fatigue in Guillain-Barré syndrome: a randomised, double blind, placebo controlled, crossover trial.* J Neurol Neurosurg Psychiatry, 2006. **77**(1): p. 61-5.
29. Greenwood, R.J., et al., *Controlled trial of plasma exchange in acute inflammatory polyradiculoneuropathy.* Lancet, 1984. **1**(8382): p. 877-9.
30. Guillain-Barré Syndrome Study Group, *Plasmapheresis and acute Guillain-Barré syndrome. The Guillain-Barré syndrome Study Group.* Neurology, 1985. **35**(8): p. 1096-104.
31. Shukla, S.K., et al., *Double blind control trial of prednisolone in Guillain-Barré syndrome - a clinical study.* Clin India, 1988. **52**: p. 128-134.
32. Bril, V., et al., *Pilot trial of immunoglobulin versus plasma exchange in patients with Guillain-Barré syndrome.* Neurology, 1996. **46**(1): p. 100-3.
33. Haupt, W.F., et al., *Sequential treatment of Guillain-Barré syndrome with extracorporeal elimination and intravenous immunoglobulin.* J Neurol Sci, 1996. **137**(2): p. 145-9.
34. Singh, N.K. and A. Gupta, *Do corticosteroids influence the disease course or mortality in Guillain-Barré' syndrome?* J Assoc Physicians India, 1996. **44**(1): p. 22-4.
35. Wollinsky, K.H., et al., *CSF filtration is an effective treatment of Guillain-Barré syndrome: a randomised clinical trial.* Neurology, 2001. **57**(5): p. 774-80.
36. Nomura, T., et al., *A randomised controlled trial comparing intravenous immunoglobulin and plasmapheresis in Guillain-Barré syndrome.* Neurol Therapeutics, 2000. **18**: p. 69-81.
37. Wang, R., et al., *Intravenous immunoglobulin in children with Guillain-Barré syndrome.* J Appl Clin Pediatr, 2001. **16**: p. 223-224.
38. Pritchard, J., et al., *A randomised controlled trial of recombinant interferon-beta 1a in Guillain-Barré syndrome.* Neurology, 2003. **61**(9): p. 1282-4.
39. Garssen, M.P., et al., *Treatment of Guillain-Barré syndrome with mycophenolate mofetil: a pilot study.* J Neurol Neurosurg Psychiatry, 2007. **78**(9): p. 1012-3.
40. French Cooperative Group on Plasma Exchange in Guillain-Barré Syndrome, *Efficiency of plasma exchange in Guillain-Barré syndrome: role of replacement fluids. French Cooperative Group on Plasma Exchange in Guillain-Barré syndrome.* Ann Neurol, 1987. **22**(6): p. 753-61.
41. Raphael, J.C., et al., *Intravenous immune globulins in patients with Guillain-Barré syndrome and contraindications to plasma exchange: 3 days versus 6 days.* J Neurol Neurosurg Psychiatry, 2001. **71**(2): p. 235-8.
42. Vermeulen, M., et al., *Intravenous immunoglobulin treatment in patients with chronic inflammatory demyelinating polyneuropathy: a double blind, placebo controlled study.* J Neurol Neurosurg Psychiatry, 1993. **56**(1): p. 36-9.
43. Hughes, R., et al., *Randomised controlled trial of intravenous immunoglobulin versus oral prednisolone in chronic inflammatory demyelinating polyradiculoneuropathy.* Ann Neurol, 2001. **50**(2): p. 195-201.
44. Hughes, R.A., et al., *Intravenous immune globulin (10% caprylate-chromatography purified) for the treatment of chronic inflammatory demyelinating polyradiculoneuropathy (ICE study): a randomised placebo-controlled trial.* Lancet Neurol, 2008. **7**(2): p. 136-44.
45. Hahn, A.F., et al., *Plasma-exchange therapy in chronic inflammatory demyelinating polyneuropathy. A double-blind, sham-controlled, cross-over study.* Brain, 1996. **119**(Pt 4): p. 1055-66.
46. Hahn, A.F., et al., *Intravenous immunoglobulin treatment in chronic inflammatory demyelinating polyneuropathy. A double-blind, placebo-controlled, cross-over study.* Brain, 1996. **119**(Pt 4): p. 1067-77.
47. Hadden, R.D., et al., *Randomised trial of interferon beta-1a in chronic inflammatory demyelinating polyradiculoneuropathy.* Neurology, 1999. **53**(1): p. 57-61.
48. Mendell, J.R., et al., *Randomised controlled trial of IVIg in untreated chronic inflammatory demyelinating polyradiculoneuropathy.* Neurology, 2001. **56**(4): p. 445-9.
49. Dyck, P.J., et al., *Combined azathioprine and prednisone in chronic inflammatory-demyelinating polyneuropathy.* Neurology, 1985. **35**(8): p. 1173-6.
50. van Doorn, P.A., et al., *High-dose intravenous immunoglobulin treatment in chronic inflammatory demyelinating polyneuropathy: a double-blind, placebo-controlled, crossover study.* Neurology, 1990. **40**(2): p. 209-12.

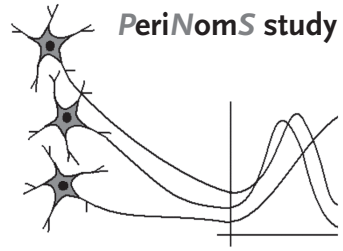
51. Van den Berg, L.H., et al., *Treatment of multifocal motor neuropathy with high dose intravenous immunoglobulins: a double blind, placebo controlled study.* J Neurol Neurosurg Psychiatry, 1995. **59**(3): p. 248-52.
52. Piepers, S., et al., *Mycophenolate mofetil as adjunctive therapy for MMN patients: a randomised, controlled trial.* Brain, 2007. **130**(Pt 8): p. 2004-10.
53. Federico, P., et al., *Multifocal motor neuropathy improved by IVIg: randomised, double-blind, placebo-controlled study.* Neurology, 2000. **55**(9): p. 1256-62.
54. Leger, J.M., et al., *Intravenous immunoglobulin therapy in multifocal motor neuropathy: a double-blind, placebo-controlled study.* Brain, 2001. **124**(Pt 1): p. 145-53.
55. Dalakas, M.C., et al., *A controlled study of intravenous immunoglobulin in demyelinating neuropathy with IgM gammopathy.* Ann Neurol, 1996. **40**(5): p. 792-5.
56. Niermeijer, J.M., et al., *Intermittent cyclophosphamide with prednisone versus placebo for polyneuropathy with IgM monoclonal gammopathy.* Neurology, 2007. **69**(1): p. 50-9.
57. Comi, G., et al., *A randomised controlled trial of intravenous immunoglobulin in IgM paraprotein associated demyelinating neuropathy.* J Neurol, 2002. **249**(10): p. 1370-7.
58. Mariette, X., et al., *A randomised clinical trial comparing interferon-alpha and intravenous immunoglobulin in polyneuropathy associated with monoclonal IgM. The IgM-associated Polyneuropathy Study Group.* J Neurol Neurosurg Psychiatry, 1997. **63**(1): p. 28-34.
59. Dyck, P.J., et al., *Plasma exchange in polyneuropathy associated with monoclonal gammopathy of undetermined significance.* N Engl J Med, 1991. **325**(21): p. 1482-6.
60. Oksenhendler, E., et al., *Plasma exchange and chlorambucil in polyneuropathy associated with monoclonal IgM gammopathy. IgM-associated Polyneuropathy Study Group.* J Neurol Neurosurg Psychiatry, 1995. **59**(3): p. 243-7.
61. Medical Research Council, *Aids to the investigation of the peripheral nervous system*. 1943, London: Her Majesty's Stationanery Office.
62. Kleyweg, R.P., F.G. van der Meche, and P.I. Schmitz, *Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barré syndrome.* Muscle Nerve, 1991. **14**(11): p. 1103-9.
63. Merkies, I.S., et al., *Fatigue in immune-mediated polyneuropathies. European Inflammatory Neuropathy Cause and Treatment (INCAT) Group.* Neurology, 1999. **53**(8): p. 1648-54.
64. Dyck, P.J., et al., *The Rochester Diabetic Neuropathy Study: design, criteria for types of neuropathy, selection bias, and reproducibility of neuropathic tests.* Neurology, 1991. **41**(6): p. 799-807.
65. Merkies, I.S., et al., *Clinimetric evaluation of a new overall disability scale in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2002. **72**(5): p. 596-601.
66. Graham, R.C. and R.A. Hughes, *A modified peripheral neuropathy scale: the Overall Neuropathy Limitations Scale.* J Neurol Neurosurg Psychiatry, 2006. **77**(8): p. 973-6.
67. Merkies, I.S., et al., *Quality of life complements traditional outcome measures in immune-mediated polyneuropathies.* Neurology, 2002. **59**(1): p. 84-91.
68. Lunn, M.P., et al., *151st ENMC international workshop: Inflammatory Neuropathy Consortium 13th-15th April 2007, Schiphol, The Netherlands.* Neuromuscul Disord, 2008. **18**(1): p. 85-9.
69. Merkies, I.S., et al., *Psychometric evaluation of a new sensory scale in immune-mediated polyneuropathies. Inflammatory Neuropathy Cause and Treatment (INCAT) Group.* Neurology, 2000. **54**(4): p. 943-9.
70. EuroQol Group, *EuroQol--a new facility for the measurement of health-related quality of life.* Health Policy, 1990. **16**(3): p. 199-208.
71. Ware Jr, J.E., et al., *SF-36 health survey Manual and interpretation guide*. 1993, The Health Institute, New England Medical Center: Boston.
72. Merkies, I.S.J., *Evaluation of scales and measurement instruments in immune-mediated polyneuropathies*. 2001, Erasmus Medical Centre: Rotterdam.
73. Merkies, I.S. and P.I. Schmitz, *Getting closer to patients: the INCAT Overall Disability Sum Score relates better to patients' own clinical judgement in immune-mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2006. **77**(8): p. 970-2.
74. Lindeboom, R., et al., *Activities of daily living instruments:optimizing scales for neurologic assessments.* Neurology, 2003. **60**(5): p. 738-42.
75. Hobart, J.C., et al., *Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations.* Lancet Neurol, 2007. **6**(12): p. 1094-105.

76. Hays, R.D., L.S. Morales, and S.P. Reise, *Item response theory and health outcomes measurement in the 21st century.* Med Care, 2000. **38**(9 Suppl): p. II28-42.
77. Bond, T.G. and C.M. Fox, *Applying the Rasch model: fundamental measurement for the human sciences*. 2001, New York: Lawrence Erlbaum Associates.
78. Bode, R.K., et al., *Issues in the development of an item bank.* Arch Phys Med Rehabil, 2003. **84**(4 Suppl 2): p. S52-60.
79. Massof, R.W., *The measurement of vision disability.* Optom Vis Sci, 2002. **79**(8): p. 516-52.
80. Andrich, D., *Controversy and the Rasch model: a characteristic of incompatible paradigms?* Med Care, 2004. **42**(1 Suppl): p. I7-16.

Chapter 1.3

*P*eripheral *N*europathy outcome measures *S*tandardisation study

*P*eri*N*om*S* study



Background

Different methods have been used to study patients with immune-mediated neuropathies.[1] The Inflammatory Neuropathy Cause and Treatment (INCAT) group, a network of European neurologists with special interest in immune-mediated neuropathies, published a series of papers and a thesis dealing with the clinimetric aspects of outcome measures used in these disorders, especially in Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyneuropathies (CIDP), and monoclonal gammopathy of undetermined significance related polyneuropathies (MGUSP).[2] To evaluate outcome in these disorders, often used outcome measures, selected by literature review, experts' opinions and patients' suggestions, were critically viewed by members of the INCAT group. Outcome measures represent different levels of outcome based on the International Classification of Functioning, Disability and Health (ICF) and on the quality of life concept, focussing more on patients' perspectives.[3, 4] The selected measures were categorised accordingly, where needed new outcome measures were created to fill the gaps. These measures were, if necessary, subsequently analysed to determine their clinimetric properties like being simple, valid, reliable, and responsive to changes over time (table 1, for further details see chapter 1.2).[5]

Table 1: Concise description of essential clinimetric properties

simple	little special training or instructions needed, not time consuming
communicable	easily understood by others
valid	measures what it supposes to measure
reliable	reproducible, consistent over time
responsive	measures relevant changes over time

Eventually, a set of clinimetrically well-evaluated outcome measures that represented various levels of outcome (impairment, disability, handicap, and quality of life) was presented, a big step forward.[2] These INCAT studies formed a base for a clinimetric-oriented workshop that was conducted in December 2004 in the Netherlands and was coordinated by the European Neuromuscular Centre (ENMC).[6] Furthermore, at a workshop on outcome measures in peripheral neuropathies held at the Peripheral Nerve Society (PNS) congress (Italy, 2005), the PNS scientific committee strongly pleaded for further standardisation. It is increasingly being recognised that consensus regarding the use of a neuropathy-specific core set of proper outcome measures is urgently needed to improve assessment during follow-up of patients with these disorders in daily clinical practice and, to evaluate their response to treatment. Therefore, the international multi-centre *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study has been conceived as part

of the Inflammatory Neuropathy Consortium (INC). INC (that succeeded INCAT) is a standing committee of the PNS, a worldwide network of neurologists who are committed to improving and advancing the investigation and treatment of immune-mediated neuropathies.

Aims

The ***P*eri*N*om*S*** study aims to expand the clinimetric knowledge on outcome measures at selected levels of outcome (pathology, impairment, activity, participation, and quality of life).

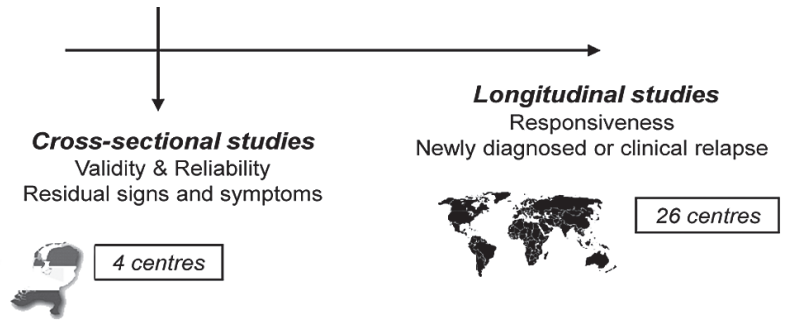
Ultimate goal: the presentation of an internationally supported standardised neuropathy-specific minimum core set of high quality outcome measures to be used in daily clinical practice and in future follow-up studies and clinical trials in patients with immune-mediated neuropathies, particularly in patients with GBS, CIDP, MMN and IgM-MGUSP.

Study design

The ***P*eri*N*om*S*** study can be divided into a cross-sectional and a longitudinal part:
The cross-sectional part, conducted in the Netherlands, focuses on examining the *reliability* and *validity* modalities of the selected outcome measures. For this part, 122 patients with an immune-mediated neuropathy and a stable clinical condition have been examined two times. During the first visit two examiners performed their scores independently and consecutively (usually within 2 hours) (inter-observer reliability). Within 2-4 weeks, one of the investigators re-examined the patient (intra-observer reliability) without having access to previous results. Additionally, 11 questionnaires about impairment, activity limitation and quality of life (test-retest) are available for further evaluation.

The longitudinal part is still being performed worldwide to obtain *responsiveness* data that may help to differentiate between comparable valid and reliable outcome measures. Therefore, newly diagnosed patients and patients with a relapse of their neuropathy (n≥140) are being examined 3 times (in MMN and MGUSP) or 5 times (in GBS and CIDP) during one-year follow up. All patients were examined at onset, 3 and 12 months. Patients with GBS and CIDP were additionally examined at 1 and 6 months because their diseases have a (sub)acute onset and usually a less indolent course compared to MMN and MGUSP.

Figure 1. ***PeriNomS*** study design



Patients

Cross-sectional patients

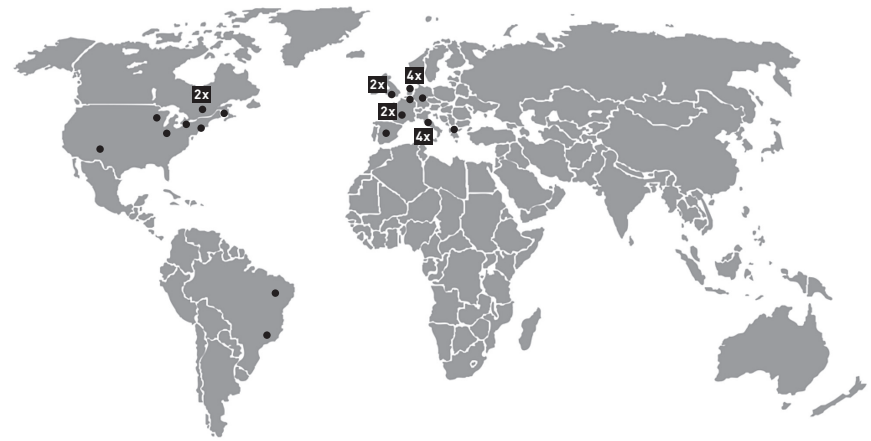
A total of 102 clinically stable patients have been included, 30 with GBS, 30 with CIDP, 20 with MGUSP, and 22 with MMN, all meeting the international criteria for their diagnosis.[7-10] A stable clinical condition was defined as: 1) an unchanged clinical functionality as declared by the patient to the best of his/her knowledge over 2 months prior to the study and during the study (stable condition at second visit compared to entry), and 2) no clear objective changes at neurological examination by the researcher when compared with recorded findings during the last 2 months before study entry (if available). All patients had residual signs and symptoms due to their illness representing a broad range of activity limitations. Patients with CIDP and MMN receiving (interval) treatment could be included as long as their clinical condition was stable (see definition above). Only IgM anti-MAG positive MGUSP patients were included. Patients with a clinical course and diagnosis compatible with CIDP and IgG MGUS were included as CIDP. Patients were excluded if having a concomitant disease or using medication like, e.g., diabetes, renal insufficiency, (prior) treatment with chemotherapy, alcohol abuse (more than 5 IU/day) that might interfere with general nervous system as well as physical functioning. Patients were recruited and investigated by trained investigators at the outpatient clinics of the university hospitals of Rotterdam, Maastricht and Utrecht, the Netherlands.

Longitudinal patients

Over 140 patients have already been included. We strive for at least 50 GBS, 50 CIDP, 20 MGUSP and 20 MMN meeting the international criteria for their diagnosis.[7-10] Patients have to be newly diagnosed or have to experience a clinical relapse or a new (multi)focal nerve lesion due to their CIDP/MMN (no medication for at least 2 months). As for the cross-sectional patients, only IgM anti-MAG positive MGUSP patients were included and IgG MGUS

patients with a clinical course and diagnosis compatible with CIDP were included as CIDP. Also, patients were excluded if having concomitant diseases or using medication like, e.g., diabetes, renal insufficiency, (prior) treatment with chemotherapy, alcohol abuse (more than 5 IU/day) that might interfere with general nervous system as well as physical functioning. Patients were recruited worldwide at 26 participating centres with expertise on peripheral neuropathies (figure 2).

Figure 2. Participating centres and list of principal investigators



Prof. D.R. Cornblath
Department of Neurology, Johns Hopkins Hospital, BALTIMORE, Maryland, USA

Prof. K.C. Gorson
Department of Neurology, Caritas St. Elizabeth's Medical Center/Tufts University School of Medicine, BOSTON, Massachusetts, USA

Prof. R.A. Lewis
Department of Neurology, Wayne State University School of Medicine, Detroit Medical Center, DETROIT, Michigan, USA

Dr. D. Walk
Department of Neurology, University of Minnesota Medical Center, MINNEAPOLIS, Minnesota, USA

Prof. M.A. Fisher
Department of Neurology, Edward Hines, Jr. VA Hospital, HINES, Illinois, USA

Dr. B.E. Smith
Department of Neurology, Mayo Clinic, Scottsdale campus, SCOTTSDALE, Arizona, USA

Prof. A.F. Hahn
Department of Neurology, London Health Science Center, LONDON, Ontario Canada

Prof. V. Bril and dr. H. Katzberg
Department of Neurology, Toronto General Hospital, TORONTO, Ontario, Canada

Dr. A.A. Barreira
Department of Neurosciences, Hospital das Clínicas de Ribeirão Preto Quarto andar, RIBEIRAO PRETO, São Paulo, Brazil

Prof. O.J.M. Nascimento
Department of Neurology, Hospital Universitário Antônio Pedro (HUAP), NITEROI, Rio de Janeiro, Brazil

Dr. M. Lunn and prof. M. Reilly
Centre for Neuromuscular disease/National Hospital for Neurology and Neurosurgery Queen Square, LONDON, United Kingdom

Dr. D. Bennett, dr. R. Hadden and prof. R.A.C. Hughes
Department of Neurology, King's College Hospital NHS Foundation Trust, LONDON, United Kingdom

Prof. E. Nobile-Orazio
Department of Neurological Sciences, Milan University, IRCCS Istituto Clinico Humanitas, MILAN, Italy

Dr. L. Padua
Institute of Neurology, Università Cattolica, ROME, Italy

G. Devigili
Neurological Unit, Ospedale dell'Angelo, VENICE, Italy

Dr. G. Lauria
Neuromuscular Diseases Unit, National Neurological Institute Carlo Besta, MILAN, Italy

Prof. M. Dalakas
Department of Pathophysiology, Unit on Neuroimmunology and Neuromuscular Diseases, ATHENS, Greece

Prof. I. Illa
Servei Neurologia, Hospital Sta. Creu i Sant Pau, BARCELONA, Spain

Prof. H-P Hartung and prof. B.C. Kieseier
Department of Neurology, Heinrich-Heine-University, DUESSELDORF, Germany

Prof. J-M Léger
Hôpital Pitié salpêtrière, Bâtiment Babinski-myologie, PARIS, France

Prof. J. Pouget
Centre de Référence des Maladies Neuromusculaires et de la SLA, Hôpital de La Timone, MARSEILLE, France

Prof. P.Y.K. van den Bergh
Centre de Référence Neuromusculaire, Cliniques universitaires St-Luc Université catholique de Louvain, BRUSSELS, Belgium

Prof. P.A. van Doorn, dr. I.S.J. Merkies and S.I. van Nes
Department of Neurology, Erasmus MC, University Medical Centre Rotterdam, ROTTERDAM, the Netherlands

Dr. C.G. Faber, dr. I.S.J. Merkies and E.K. Vanhoutte
Department of Neurology, Maastricht University Medical Centre, MAASTRICHT, the Netherlands

Prof. L.H. van den Berg, dr. N.C. Notermans and dr. W.L. van der Pol
Rudolph Magnus Institute, University Medical Centre Utrecht, UTRECHT, the Netherlands

Prof. I.N. van Schaik, dr. A. van der Kooi and prof. M. de Visser
Department of Neurology, Academic Medical Centre, AMSTERDAM, the Netherlands

Selected outcome measures

Table 2 shows the selected outcome measures and also notes the ones used for the studies in this thesis.

Table 2. Selected outcome measures for the ***PeriNomS*** study

Level of outcome	Outcome measure	Assessment frequency		
		Cross-sectional group	Longitudinal group	
			GBS CIDP	MGUSP MMN
Pathology	IENF density	1x□	2x	2x□
Impairment	Jamar dynamometer*	3x	5x	3x
	Vigorimeter*	3x	5x	3x
	MRC sumscore*	3x	5x	3x
	NIS motor subset	3x	5x	3x
	INCAT sensory sumscore*	3x□	5x	3x□
	NIS sensory subset*	3x□	5x	3x□
	EMG	1x	2x	2x
	mD-COMPASS	2x□	5x	3x□
	11-point PI-NRS	2x□	5x	3x□
	VAS-pain	2x□	5x	3x□
Activity and participation	ODSS*	3x	5x	3x
	ONLS*	3x	5x	3x
	R-ODS/ R-ODS-MMN*	2x	5x	3x
Quality of life	Vickrey's scale	2x	5x	3x
	SF-36	2x	5x	3x
	EuroQoL-5D	2x	5x	3x
	SIP	2x	5x	3x
	NHP	2x	5x	3x
	WHO-QoL bref	2x	5x	3x
	VAS-QoL	2x	5x	3x
Composite levels	PPCM	-	5x	3x
	clinical judgment score*	-	5x	3x

Legend to table 2. *= outcome measures used for studies in this thesis. □ = not in patients with MMN. IENF= intraepidermal nerve fibre, MRC= Medical Research Council, NIS= neuropathy impairment score, INCAT= Inflammatory Neuropathy Cause and Treatment group, EMG= electromyography, mD-COMPASS= modified Dutch composite autonomic symptom scale, PI-NRS= pain intensity numerical scale, VAS= visual analogue scale, ODSS= overall disability sumscore, ONLS= overall neuropathy limitations scale, R-ODS= Rash-built overall disability scale, R-ODS MMN= Rash-built overall disability scale for MMN, Vickrey's scale is a peripheral neuropathy quality of life instrument, SF-36= short form 36-item health survey, EuroQoL-5D= EuroQoL group-5D scale, SIP= sickness impact profile, NHP= Nottingham health profile, WHO-QoL brief = short form of the WHO quality of life scale, QoL= quality of life, PPCM= personal patient-centred measures

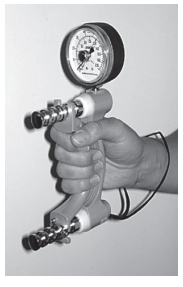
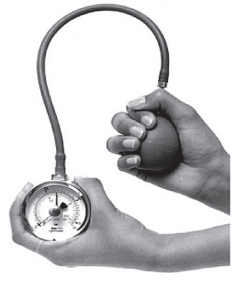
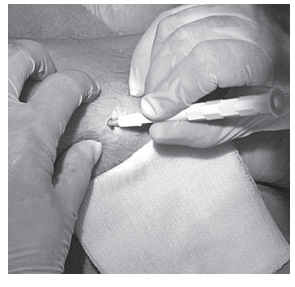
The participating investigators were trained in 2007 at the PNS meeting (Utah, USA) aiming to standardise the assessment procedures for all scales as part of this study. In addition, participants received a comprehensive research manual that integrated a thorough description and pictures illustrating how to assess the various outcome measures. In the following, a short description of the selected outcome measures:

At the pathology level of outcome:

- *Intraepidermal nerve fibre (IENF) density*
  IENF density is the number of individual nerve fibres per unit area crossing the dermal-epidermal junction. To determine IENF density according to European guidelines, a skin biopsy, a punch biopsy of 3 mm, 10 cm above the right lateral malleolus was performed after local anaesthesia with 1% lidocaine.[11] Skin biopsy is a routine diagnostic procedure in several centres, but determination of IENF density is not. Therefore, skin biopsies were only performed in centres with sufficient expertise to determine IENF density (the pathology laboratories in Milan and Maastricht). As a result, biopsies were taken once in all cross-sectional patients and twice in longitudinal Dutch and Italian patients who gave informed consent. No biopsies were taken from MMN patients.

At the impairment level of outcome:

- *Vigorimeter and Jamar dynamometer*
  Both tools are designed to assess grip strength and were assessed at each visit in GBS, CIDP, MGUSP and MMN patients.[2, 12, 13] The Jamar dynamometer quantifies isometric force in pounds (range 0-200). The Vigorimeter quantifies grip strength on a manometer after squeezing a rubber bulb, and is scored in kilopascal (range 0-160). All patients were asked to judge both the Vigorimeter and the Jamar regarding their comfortability in assessing grip strength when compared to each other (1 = great preference for Vigorimeter; 2 = little preference for Vigorimeter; 3 = no preference for any of the two tools; 4 = little preference for Jamar; 5 = great preference for Jamar).

Jamar dynamometer	Vigorimeter	Punch biopsy
		

- *Medical Research Council (MRC) sumscore*
  At each visit the MRC sumscore was assessed, a summation of MRC grades given in full numbers (0= no movement, no contraction, 1= visible contraction without movement, 2= movement but only with gravity eliminated, 3= movement against gravity, 4= movement against resistance, but weaker than normal, 5= normal strength) of the arm abductors, elbow flexors, wrist extensors, hip flexors, knee extensors and foot dorsal flexors at both sides and thereby, ranging from 0 (total paralysis) to 60 (normal strength).[14] An expanded MRC sumscore ranging from 0 (total paralysis) to 140 (normal strength) was used in MMN to cover its asymmetrical presentation and its predominantly affected distal arm muscles. Based on expert opinion the following fourteen muscle pairs, including the ten previously described as the most affected ones, were selected to compose this summation: shoulder abductors, elbow extensors and flexors, wrist extensors and flexors, finger extensors and flexors, thumb abductors, hypothenar abductors, hip flexors, knee extensors and flexors, ankle dorsiflexors and ankle plantar flexors.[15] For each muscle group standardised joint/limb positions, point of counterforce and some MRC grading scores were defined and illustrated in the provided research manual.
- *Neuropathy impairment score (NIS)*
  This scale consists of a motor and a sensory subscale.[16,17] The NIS motor subscale ranges from 0 (no motor deficit) to 48 (most severe motor deficit) and is composed by grading each of the six assessed muscle groups (arm abductors, forearm flexors, wrist extensors, hip flexors, knee extensors and foot dorsal flexors) at both sides as follows: 0 = normal, 1 = 25% weak, 2 = 50% weak, 3 = 75% weak, 3.25 = move against gravity, 3.5 = movement, gravity eliminated, 3.75 = muscle flicker, no movement, 4 = paralysis. An MMN-specific motor subscale, ranging from 0 (no motor deficit) to 112 (most severe motor deficit) was composed by the summation of the following fourteen muscle pairs: shoulder abductors, elbow extensors and flexors, wrist extensors and flexors, finger extensors and flexors, thumb abductors, hypothenar abductors, hip flexors, knee extensors and flexors, ankle dorsiflexors and ankle plantar flexors using the above mentioned NIS grading system.
  The NIS sensory subscale, ranging from 0 (no sensory deficit) to 32 (most severe sensory deficit), is composed by the summation of the following sensation qualities: touch pressure, pinprick, vibration, and joint position sense. Touch pressure, pinprick and vibration sensation are tested on the dorsal surface of the index finger and great toe, at the base of the nail of the terminal phalanx at both sides. Touch pressure is assessed with long fibre cotton wool, pinprick is assessed with straight pins and vibration sensation is tested with a 165 Hz tuning fork. Joint position is tested by moving the terminal phalanx of the index finger and great toe at both sides. Each quality is scored according to the

following scoring system: 0 = normal, 1 = decreased and 2 = absent sensation. For both subscales the examiner scores deficits by what he (she) considers to be normal considering test, anatomical site, age, gender, weight, and physical fitness. Detailed information about the NIS (motor and sensory subscale) is also available by video-instruction. Each visit both subscales were assessed, except for the sensory subscale in MMN patients.

Table 3. Scoring table for the sensory subset of the NIS - sites of examination with corresponding grades. Tick the box that applies.

Scoring sensation **NA= not applicable; 0 = normal; 1 = decreased; 2 = absent**								
	RIGHT				**LEFT**			
Sensation – index finger	**NA**	**0**	**1**	**2**	**NA**	**0**	**1**	**2**
Touch pressure	O	O	O	O	O	O	O	O
Pinprick	O	O	O	O	O	O	O	O
Vibration	O	O	O	O	O	O	O	O
Joint position	O	O	O	O	O	O	O	O
Sensation – great toe	**NA**	**0**	**1**	**2**	**NA**	**0**	**1**	**2**
Touch pressure	O	O	O	O	O	O	O	O
Pinprick	O	O	O	O	O	O	O	O
Vibration	O	O	O	O	O	O	O	O
Joint position	O	O	O	O	O	O	O	O

- *(modified) INCAT sensory sumscore ((m)ISS)*
  The ISS is composed by the summation of the following sensation qualities: pinprick and vibration sense in the arms and legs from distal to proximal recording the highest extension of dysfunction, plus two-point discrimination at the index finger. Joint position and light touch qualities however are not incorporated in the original ISS, making this scale less comprehensive and perhaps less responsive.[6] Furthermore two-point discrimination categories were arbitrarily defined.[18] Therefore, a modified ISS was constructed incorporating light touch and joint position sense and grading the collected two-point discrimination values. Light touch was examined with a cotton wool stick and pin prick using a broken cotton wool stick. Vibration sense was examined using a Rydel Seiffer 64 Hz graduated tuning fork and graded comparing the findings with published normative data.[19] Joint position was assessed in a standardised manner based on previous recommendations, with predefined joint/limb starting positions

and examiner's positions to fixate and initiate movement as illustrated in the provided research manual.[20, 21] The above mentioned qualities were graded as normal (grade 0) or disturbed; at the index finger or hallux (grade 1); at the wrist or ankle (grade 2), at the elbow or knee (grade 3), at shoulder or hip joint (grade 4). Two-point discrimination was assessed at the right index finger only using a sliding aesthesiometer and graded as normal (grade 0) or disturbed (grade 1) based on age-dependent normative values now available (see also chapter 2.1).[22] The modified ISS, ranging from 0 (no sensory deficit) to 33 (most severe sensory deficit), was assessed at each visit in patients with GBS, CIDP, and MGUSP.

Table 4. Modified INCAT sensory sumscore: sites of examination with corresponding grades.

Sensation		**Normal**	**Abnormal**			
	Grade	**0**	**1**	**2**	**3**	**4**
Pinprick	**Arms**	at index finger	at index finger	at ulnar styloid process	at medial humerus epicondyle	at acromio-clavicular joint
	Legs	at hallux	at hallux	at medial malleolus	at patella	at anterior superior iliac spine
Light touch	**Arms**	at index finger	at index finger	at ulnar styloid process	at medial humerus epicondyle	at acromio-clavicular joint
	Legs	at hallux	at hallux	at medial malleolus	at patella	at anterior superior iliac spine
Vibration sense	**Arms**	at index finger	at index finger	at ulnar styloid process	at medial humerus epicondyle	at acromio-clavicular joint
	Legs	at hallux	at hallux	at medial malleolus	at patella	at anterior superior iliac spine
Joint position	**Arms**	DIP joint index finger	DIP joint index finger	at wrist	at elbow	at shoulder joint
	Legs	DIP joint hallux	DIP joint hallux	at ankle	at knee	at hip joint
Two-point discrimination	**Index finger**	at index finger	at index finger			

Legend to table 4. DIP= distal interphalangeal

- *Electromyography (EMG)*
  If informed consent was given a routine EMG was performed once in the cross-sectional patients and twice, at entry and after one year follow-up, in all longitudinally studied patients. A standardised set of electrophysiological measurements was performed in the right arm and leg (in the left side if not possible due to e.g., injuries or deformities) using conventional techniques with surface electrode recordings after limb warming-up procedure.[23, 24] Tests included assessment of motor nerves: median (two-point stimulation: wrist and distal to elbow), ulnar (three-point stimulation: wrist, distal and proximal to elbow), peroneal (three-point stimulation: ankle, fibular head, popliteal fossa), and tibial (two-point stimulation: ankle, popliteal fossa) to determine the compound muscle action potential following distal and proximal stimulation, terminal latencies, conduction velocities, and F-response latencies. Sensory nerves were also assessed: the median (II), ulnar (V) and sural nerves were stimulated antidromically to determine evoked sensory nerve action potential amplitudes, distal latencies and conduction velocities. Needle electromyography of first dorsal interosseous, vastus medialis and anterior tibial muscles were carried out to evaluate the presence of fibrillation potentials and positive sharp waves.
- *Modified Dutch composite autonomic symptom scale (mD-COMPASS)*
  COMPASS is an instrument to assess autonomic symptoms. It correlates well with the CASS, a composite autonomic scoring scale that encompasses several autonomic function tests.[25] The English version of the COMPASS was translated into Dutch, according to the international guidelines.[26] In the modified Dutch version, we added questions concerning female sexual dysfunction. The assessment of the mD-COMPASS starts the moment patients commence with mobilisation, covering at least 6 hours out of their bed or directly at hospital admission if patients are not bed bound. Except for MMN patients, all patients were asked to fill in the mD-COMPASS at each visit.
- *Visual analogue scale for pain (VAS-pain) and 11-point pain intensity numerical scale (11 point PI-NRS)*
  Except for MMN patients, all patients were instructed to complete these pain scales at each visit conform the published procedures.[27, 28]

At the activity limitation level of outcome:

- *Overall disability sumscore (ODSS), overall neuropathy limitation scale (ONLS)*
  Both the ODSS and the ONLS are composed of a summation of an arm disability scale (range: 0-5) and a leg disability scale (range: 0-7). The examiner questions and observes the patient in order to determine an arm grade and a leg grade. Each visit these activity limitation scales were completed according to the published procedures for each.[2, 29]

- *Rasch-built overall disability scale (R-ODS)*
  New activity and participation limitation scales were constructed reflecting the specific limitations of patients with immune-mediated neuropathies using the Rasch method (see also Chapter 3.3 and 3.4).[30] This method enables the use of sumscores by creating interval scales revealing the real difference in ability levels between patients and within patients in time. Both, the R-ODS for patients with GBS, CIDP and MGUSP and the R-ODS specifically designed for MMN patients (R-ODS-MMN), were assessed at each visit.

At the quality of life level of outcome:

- *Vickrey's scale, short form 36-item health survey (SF-36), EuroQoL-5D, sickness impact profile (SIP), Nottingham health profile (NHP), and short form of the WHO quality of life scale (WHOQoLbref)*
  The Vickrey's scale is a health related quality of life measure specific for patients with peripheral neuropathies.[31] The other health-related scales concern generic quality of life measures.[32-43] At each visit, all patients were asked to complete all quality of life questionnaires according to the published procedures by each one of them. Furthermore, at each visit patients were asked to tick a box on the quality of life visual analogue scale (QoL-VAS).

At the composite level of outcome:

- *Clinical judgment score (ClinJSc)*
  At each visit, patients were requested to judge whether their clinical condition strongly deteriorated (coded 1), slightly deteriorated (coded 2), remained stable (coded 3), slightly improved (coded 4), or strongly improved (coded 5) when compared with last visit (= defined as '*clinical judgment score*'). At study entry, patients reflected their clinical condition against their physical status within the two weeks before the start of the study.
- *Personal patient-centred measures (PPCM)*
  These are based on the Canadian occupational performance measure (COPM) concept of self-reported activities using a semi-structured interview undertaking the following steps: the researcher helps the patient to identify problem areas in daily activities and social participation. A problem exists when a patient cannot do, doesn't do or isn't satisfied with how he is doing his occupation or a particular task. By focusing on the patients' own roles and environments we tried to identify areas that are relevant to them. After identifying problem areas, patients were asked to order the selected problems by their priorities (from most important to less important). Subsequently, the 5 most important problem areas to the patient were selected for follow-up and scored at each visit. These items are scored on a 5-point rating scale (1 meaning 'not able to do' or 'not satisfied at all' to 5 meaning 'able to do it extremely well' or 'extremely satisfied') depending on the personal performance and satisfaction by the patient.[44, 45]

Specific *PeriNomS* study objectives within the scope of this thesis

For cross-sectional part

At the impairment level of outcome:

- *Reliability* and *validity* studies in patients with GBS, CIDP, and MGUSP:
  - Comparison between the hand-held Vigorimeter and the Jamar dynamometer.
  - Comparison between (modified) INCAT sensory sumscore and NIS sensory subset.
- *Reliability* and *validity* studies in patients with MMN:
  - Comparison between the hand-held Vigorimeter and the Jamar dynamometer.

At the activity and participation level of outcome:

- *Reliability* and *validity* studies in all patients group:
  - Comparison between the newly devised Rasch-built overall disability scores (R-ODS) – one for patients with GBS/CIDP/MGUSP and one for patients with MMN – and the ODSS and ONLS separately.

For longitudinal part

At the impairment level of outcome:

- *Responsiveness* studies in patients with GBS, CIDP, and MGUSP:
  - Comparison between (modified) INCAT sensory sumscore and NIS sensory subset
  - Comparison between the hand-held Vigorimeter and the Jamar dynamometer
- *Responsiveness* studies in patients with MMN:
  - Comparison between the hand-held Vigorimeter and the Jamar dynamometer.

At the activity and participation level of outcome:

- *Responsiveness* studies in all patients group:
  - Comparison between the newly devised Rasch-built overall disability scale (R-ODS) – one for patients with GBS/CIDP/MGUSP and one for patients with MMN – and the ODSS and ONLS separately.

References

1. van Nes, S.I., C.G. Faber, and I.S. Merkies, *Outcome measures in immune-mediated neuropathies: the need to standardize their use and to understand the clinimetric essentials.* J Peripher Nerv Syst, 2008. **13**(2): p. 136-47.
2. Merkies, I.S.J., *Evaluation of scales and measurement instruments in immune-mediated polyneuropathies. Thesis*, in *Neurology*. 2001, Erasmus Medical Center: Rotterdam.
3. World Health Organization, *International classification of impairments, disabilities, and handicaps*. 2001: Geneva.
4. Aaronson, N.K., *Quality of life: what is it? How should it be measured?* Oncology (Williston Park), 1988. **2**(5): p. 69-76, 64.
5. Feinstein, A.R., *Clinimetrics*. 1987, New Haven and London: Yale University Press.
6. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
7. Asbury, A.K. and D.R. Cornblath, *Assessment of current diagnostic criteria for Guillain-Barre syndrome.* Ann Neurol, 1990. **27 Suppl**: p. S21-4.
8. *Research criteria for diagnosis of chronic inflammatory demyelinating polyneuropathy (CIDP). Report from an Ad Hoc Subcommittee of the American Academy of Neurology AIDS Task Force.* Neurology, 1991. **41**(5): p. 617-8.
9. Miescher, G.C. and A.J. Steck, *Paraproteinaemic neuropathies.* Baillieres Clin Neurol, 1996. **5**(1): p. 219-32.
10. van Schaik, I.N., et al., *European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of multifocal motor neuropathy.* Eur J Neurol, 2006. **13**(8): p. 802-8.
11. Lauria, G., et al., *EFNS guidelines on the use of skin biopsy in the diagnosis of peripheral neuropathy.* Eur J Neurol, 2005. **12**(10): p. 747-58.
12. Mathiowetz, V., *Grip and pinch strength measurements*, in *Muscle strength testing. Instrumented and non-instrumented systems*, L.R. Amundsen, Editor. 1990, Churchill Livingstone: New York. p. 163-177.
13. Desrosiers, J., et al., *Comparison of the Jamar dynamometer and the Martin vigorimeter for grip strength measurements in a healthy elderly population.* Scand J Rehabil Med, 1995. **27**(3): p. 137-43.
14. Kleyweg, R.P., F.G. van der Meche, and P.I. Schmitz, *Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barre syndrome.* Muscle Nerve, 1991. **14**(11): p. 1103-9.
15. Van Asseldonk, J.T., et al., *Demyelination and axonal loss in multifocal motor neuropathy: distribution and relation to weakness.* Brain, 2003. **126**(Pt 1): p. 186-98.
16. Dyck, P.J., et al. *The Rochester Diabetic Neuropathy Study: design, criteria for types of neuropathy, selection bias, and reproducibility of neuropathic tests.* Neurology, 1991. **41**(6): p. 799-807.
17. Dyck, P.J., et al. *The Rochester Diabetic Neuropathy Study: reassessment of tests and criteria for diagnosis and staged severity.* Neurology, 1992. **42**(6): p. 1164-70.
18. Dyck, P.J., R.A.C. Hughes, and P.C. O'Brien, *Quantitating overall neuropathy symptomes, and outcomes*, in *Peripheral Neuropathy*, P.J. Dyck and P.K. Thomas, Editors. 2005, WB Saunders Company: Philadelphia.
19. Martina, I.S., et al., *Measuring vibration threshold with a graduated tuning fork in normal aging and in patients with polyneuropathy. European Inflammatory Neuropathy Cause and Treatment (INCAT) group.* J Neurol Neurosurg Psychiatry, 1998. **65**(5): p. 743-7.
20. Gilman, S., *Joint position sense and vibration sense: anatomical organisation and assessment.* J Neurol Neurosurg Psychiatry, 2002. **73**(5): p. 473-7.
21. Stillman, B., *Thesis: an investgation of the clinical assessment of joint position sense*, in *School of physiotherapy*. 2000: Victoria, Australia.
22. van Nes, S.I., et al., *Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies.* J Neurol Neurosurg Psychiatry, 2008. **79**(7): p. 832-4.
23. Franssen, H. and G.H. Wieneke, *Nerve conduction and temperature: necessary warming time.* Muscle Nerve, 1994. **17**(3): p. 336-44.

24. Franssen, H., N.C. Notermans, and G.H. Wieneke, *The influence of temperature on nerve conduction in patients with chronic axonal polyneuropathy.* Clin Neurophysiol, 1999. **110**(5): p. 933-940.
25. Suarez, G.A., et al., *The Autonomic Symptom Profile: a new instrument to assess autonomic symptoms.* Neurology, 1999. **52**(3): p. 523-8.
26. Streiner, D.L. and G.R. Norman, *Health measurement scales. A practical guide to their development and use*. 1998, New York: Oxford University Press, 2nd ed.
27. Maxwell, C., *Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment.* Br J Clin Pharmacol, 1978. **6**(1): p. 15-24.
28. Farrar, J.T., et al., *Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale.* Pain, 2001. **94**(2): p. 149-58.
29. Graham, R.C. and R.A. Hughes. *A modified peripheral neuropathy scale: the Overall Neuropathy Limitations Scale.* J Neurol Neurosurg Psychiatry, 2006. **77**(8): p. 973-6.
30. van Nes, S.I., et al., *Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies.* Neurology, 2011. **76**(4): p. 337-45.
31. Vickrey, B.G., R.D. Hays, and M. Beckstrand, *Development of a health-related quality of life measure for peripheral neuropathy.* Neurorehabil Neural Repair, 2000. **14**(2): p. 93-104.
32. Bergner, M., et al., *The Sickness Impact Profile: development and final revision of a health status measure.* Med Care, 1981. **19**(8): p. 787-805.
33. Jacobs, H.M., et al., *[The sickness impact profile; results of an evaluation study of the Dutch version].* Ned Tijdschr Geneeskd, 1990. **134**(40): p. 1950-4.
34. Aaronson, N.K., et al., *Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations.* J Clin Epidemiol, 1998. **51**(11): p. 1055-68.
35. Erdman, R.A., et al., *The Dutch version of the Nottingham Health Profile: investigations of psychometric aspects.* Psychol Rep, 1993. **72**(3 Pt 1): p. 1027-35.
36. Guyatt, G.H., D.H. Feeny, and D.L. Patrick, *Measuring health-related quality of life.* Ann Intern Med, 1993. **118**(8): p. 622-9.
37. Hyland, M.E., *A brief guide to the selection of quality of life instrument.* Health Qual Life Outcomes, 2003. **1**: p. 24.
38. Sprangers, M.A., et al., *Assessing meaningful change in quality of life over time: a users' guide for clinicians.* Mayo Clin Proc, 2002. **77**(6): p. 561-71.
39. Saxena, S., D. Carlson, and R. Billington, *The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research.* Qual Life Res, 2001. **10**(8): p. 711-21.
40. Trompenaars, F.J., et al., *Content validity, construct validity, and reliability of the WHOQOL-Bref in a population of Dutch adult psychiatric outpatients.* Qual Life Res, 2005. **14**(1): p. 151-60.
41. *EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group.* Health Policy, 1990. **16**(3): p. 199-208.
42. Ware Jr, J.E., et al., *User's manual for the SF-36v2 Health survey*. Vol. 2nd edition. 2007, Lincoln: QualityMetric incorporated.
43. Ware Jr, J.E., et al., *SF-36 health survey Manual and interpretation guide*. 1993, The Health Institute, New England Medical Center: Boston.
44. Law, M., et al., *The Canadian occupational performance measure: an outcome measure for occupational therapy.* Can J Occup Ther, 1990. **57**(2): p. 82-7.
45. Law, M., et al., *Pilot testing of the Canadian Occupational Performance Measure: clinical and measurement issues.* Can J Occup Ther, 1994. **61**(4): p. 191-7.

Chapter 1.4

Objectives and outline of this thesis

Objectives and outline of this thesis

This thesis focuses on improving and standardising assessment of patients with immune-mediated neuropathies by;

1. **providing normative values to improve scoring systems of existing scales and tools**
   - revised normative values for the two-point discriminator to define (ab)normal for this quality in the modified INCAT sensory sumscore (*chapter 2.1*)
   - revised normative values for measuring grip strength with the Jamar dynamometer (*chapter 2.2*)

2. **introducing Rasch analyses to improve existing scales and to create interval scales**
   - improving the fatigue severity scale by using Rasch *(chapter 3.2)*
   - creating an activity and participation limitation scale for GBS, CIDP and MGUSP *(chapter 3.3)* and for MMN *(chapter 3.4)* by using Rasch

3. **selecting outcome measures based on comparative validity, reliability and responsiveness studies**
   - comparison of the Jamar dynamometer and the Vigorimeter to assess grip strength *(chapter 4.1)*.
   - comparison of the sensory subset of the neuropathy impairment score (NISs) and the modified INCAT sensory sumscore (mISS) *(chapter 4.2)*

4. **introducing the concept of minimum clinically important difference (MCID) to define a responder**
   - using MCID cut offs to define clinically relevant change *(chapter 5.2)*
   - dynamic MCID based on individual standard errors *(chapter 5.3)*

OBJECTIVES OF THIS THESIS:

- to construct a neuropathy-specific core set of high quality outcome measures for future immune-mediated neuropathy studies and daily clinical practice (aim of the ***P*eri*N*om*S*** study)
- to contribute to the shift from ordinal outcome measures to measures with a linear construct using Rasch analysis
- to demonstrate the challenges of defining a responder using the concept of MCID

Chapter 2

Normative value studies

Chapter 2.1

Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies

Sonja I. van Nes[1], Catharina G. Faber[2], Rinske M.T.P. Hamers[2], Oliver Harschnitz[2], Mayienne Bakkers[2], Mieke C.E. Hermans[2], Ron J. Meijer[3], Pieter A. van Doorn[1], and Ingemar S.J. Merkies[1,2,3], on behalf of the PeriNomS study group

[1]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of Neurology, University hospital Maastricht, Maastricht, the Netherlands
[3]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands

J Neurol Neurosurg Psychiatry 2008 Jul;79(7):832-34

Abstract

Objectives: To revise the static and dynamic normative values for the two-point discrimination test and to examine its applicability and validity in patients with a polyneuropathy.

Methods: Two-point discrimination threshold values were assessed in 427 healthy controls and 99 patients mildly affected by a polyneuropathy. The controls were divided into seven age groups ranging from 20-29, 30-39,..., up to 80 years and older; each group consisting of at least 30 men and 30 women. Two-point discrimination examination took place under standardised conditions on the index finger. Correlation studies were performed between the scores obtained and the values derived from the Weinstein enhanced sensory test (WEST) and the arm grade of the overall disability sumscore (ODSS) in the patients' group (validity studies). Finally, the sensitivity to detect patients mildly affected by a polyneuropathy was evaluated for static and dynamic assessment.

Results: There was a significant age dependent increase in the two-point discrimination values. No significant gender difference was found. The dynamic threshold values were lower than the static scores. The two-point discrimination values obtained correlated significantly with the arm grade of the ODSS (static values: r=0.33, p=0.04; dynamic values: r=0.37, p=0.02) and the scores of the WEST in patients (static values: r=0.58, p=0.0001; dynamic values: r=0.55, p=0.0002). The sensitivity for the static and dynamic threshold values was 28% and 33%, respectively.

Conclusion: This study provides age-related normative two-point discrimination values using a two-point discriminator (an aesthesiometer). This easily applicable instrument could be used as part of a more extensive neurological sensory examination.

Introduction

The assessment of sensory deficit in clinical practice tends to be less obvious and more prone to subjective interpretation than the evaluation of motor dysfunction. This may explain the diversity in sensory scales used in polyneuropathy studies. To overcome these limitations and to strive for uniformity, the Inflammatory Neuropathy Cause and Treatment (INCAT) group introduced and clinimetrically evaluated the INCAT sensory sumscore (ISS).[1] The ISS comprises vibration and pinprick sense plus a two-point discrimination value using a sliding aesthesiometer. On the basis of experts' opinion, this instrument was assessed in a 'static' manner at the index finger. The scoring system was chosen arbitrarily, because the reference values in the available studies turned out to be rather conflicting. In these studies different groups were investigated with poor or no information on stratification for age and gender using different methodologies.[2-6] To date, there is also no consensus about the assessment of two-point discrimination quality in a 'static' or 'dynamic' manner. On the basis of these shortcomings and striving for clinically applicable reference values, we examined a large cohort of healthy individuals to obtain normative values. Moreover, the validity and sensitivity of two-point discrimination assessment were investigated in patients mildly affected by a polyneuropathy.

Participants and methods

Healthy controls

A total of 427 healthy controls were included. They were recruited from hospital personnel, relatives and friends of patients, and from homes for the elderly. Healthy individuals were stratified for age and gender, forming seven age decade groups (20-29, 30-39, ..., up to ≥ 80 years) consisting of at least 30 men and 30 women each. Eligibility criteria were: lucid consciousness and no history of mental or psychological illness, no history of alcohol misuse, no usage of drugs or history of diseases that may cause sensory deficit or influences cooperation, independence in activities of daily living, no sensory symptoms in hands or feet (e.g., burning, tingling, numbness), absence of any impairment affecting upper limbs (e.g., joint problems, deformities, muscle ache), present tendon reflexes and normal sensory modalities (pain, touch and vibration sense) at examination. Pain was tested with disposable pins, touch with cotton wool sticks and vibration with a Rydel-Seiffer tuning fork using the published reference values.[7]

Patients with a polyneuropathy

Ninety-nine patients were recruited from the university hospital of Maastricht outpatients' databank. They were all able to walk independently. The

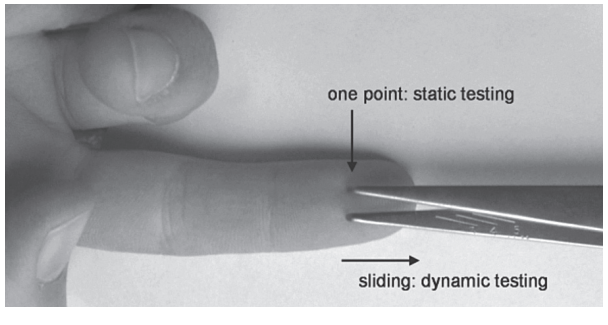
aetiologies were diabetes mellitus (n=23), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) (n=23), Guillain-Barré syndrome (n=14), uraemia (n=13), drug-induced (n=6), hereditary motor and sensory neuropathy (HMSN) type I (n=5), and vitamin B12 deficiency (n=2). In 13 cases no cause was determined.

Assessment tools and scale

- A *two-point discriminator* with adjustable points (an aesthesiometer) was used (figure 1). The exact measurable distance can be read on one arm of the instrument (range 3 mm-10 cm).
- The *Weinstein enhanced sensory test (WEST)* delivers calibrated forces to intact skin by use of five force-calibrated monofilaments.[8]
- The *arm grade of the overall disability sumscore (ODSS)* was selected, as it captures upper limb function through daily activities (e.g., dressing upper part of the body, using knife and fork).[9] Good clinimetric properties have been demonstrated for the WEST and the ODSS.

Figure 1. Two-point discriminator (aesthesiometer)

Figure 2. Static and dynamic two-point discrimination assessment



Test procedure

All participants gave informed consent before the study. The medical ethics committee of the university of Maastricht approved the study. Assessments were performed in a random order. For two-point discrimination assessment, the ends of the arms of the aesthesiometer were placed simultaneously on the right index finger. The aesthesiometer was held at its hinge (turning point) while the ends of the arms were rested gently on the skin without application of any pressure, only the weight of the instrument (30 g, length of arms 12,4 cm). The distance between the two ends was varied. The initial distance started at >2 cm (which was enough for almost all controls to detect the two points), gradually descending until the participant could not differentiate between the two points. First an example of one and of two points was given. Subsequently, to obtain a threshold value, a subject had to differentiate correctly between two points at a given distance 7 out of 10 times. Catch trials were randomly applied to enhance measurements. Participants were examined in a standardised position; sitting with the forearm in supination, resting comfortably on an armchair or desk, and with eyes closed. Static examination was performed by applying the ends of the arms of the aesthesiometer to one point at the distal phalanx. For dynamic examination, the ends of the arms were gently moved from the proximal to the distal end of the distal phalanx, over a distance of approximately 1 cm. Contact with the skin was maintained while the ends were moved perpendicular to the gap between the two-points. Three data-collecting series were performed for both static and dynamic assessments. In cases of doubt, the series were expanded to a maximum of five. The WEST was applied according to the instructions given by the manufacturer (Connecticut Bioinstruments, Riverdale, New York, USA).[8]

Statistical analysis

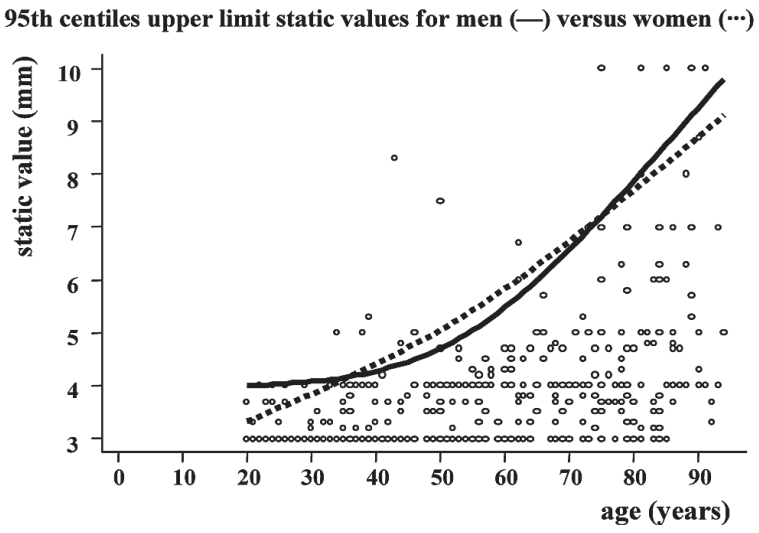
Static and dynamic reference values were calculated in healthy controls for the two-point discrimination test (95^{th} centile values, corresponding to a chosen specificity of 95%), depending primarily on age and sex, using quantile regression analyses with restricted cubic spline functions on age. The limits obtained were estimated and further used to determine the sensitivity for the static and dynamic approach. A two-point discrimination score was considered to be abnormal if the corresponding value was above the 95^{th} centile limit.

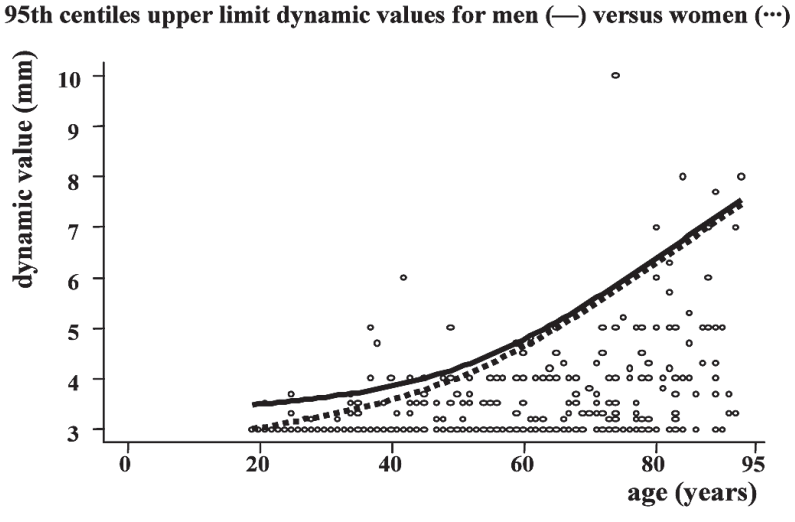
Construct validity for the aesthesiometer was obtained by correlating the two-point discrimination values with the values of the WEST and the arm grade of the ODSS in patients (Spearman rank test). All analyses were performed using Stata 7.0 for Windows 2000. A value of $p<0.05$ was considered statistically significant.

Results

A total of 427 healthy controls were included (all age groups having 30 men and 30 women, except age group 20-29 years (31 men and 31 women) and age group ≥ 80 years (30 men and 35 women). There was an age-dependent increase in the acquired static and dynamic two-point discrimination values. The corresponding graphics show these increase reflecting the decline in two-point discrimination sense with aging (figure 3).

Figure 3. results of static and dynamic measurement with the aesthesiometer





Threshold values obtained from the dynamic assessment turned out to be lower than those acquired from static examination (figure 3). The 95% centile normative static and dynamic scores were independent of gender and were subsequently translated for clinical use (table 1).

Table 1. Normative static and dynamic two-point discrimination values

Age (years)	Static assessment (mm)	Dynamic assessment (mm)
20-39	4.0	3.5
40-49	4.5	4.0
50-59	5.0	4.0
60-69	6.0	5.0
70-79	7.0	6.0
≥ 80	8.5	6.5

Legend to table 1. Normative values are 95th centile values per age span. mm=millimetres

A total of 37 women and 62 men with a polyneuropathy were included (mean (SD) age 58.9 (13,6), range 26-87 years). No disability in upper limb functioning was found in 34.1% of the patients. Minor (ODSS arm grade 1: 9.8%), moderate (grade 2: 48.8%) and severe (grade 3: 7.3%) signs were observed in the remaining patients. A significant correlation was found between the two-point discrimination values obtained and the ODSS arm grade (Spearman rank test: static values versus arm grade: r=0.33, p=0.04, dynamic values versus arm grade: r=0.37, p=0.02), and the WEST scores in these patients (Spearman rank test: static values versus WEST: r=0.58, p=0.0001; dynamic values versus WEST: r=0.55, p=0.0002). The sensitivity of the aesthesiometer (ability to detect a patient) was 28% and 33% when assessed in a static or dynamic manner, respectively.

Discussion

This study provides revised static and dynamic normative two-point discrimination threshold values. These values increase with age, but are gender independent. This is in line with earlier reports, although different devices and assessment methods were used.[2-5] Various factors may contribute to the increase in two-point discrimination values with increasing age: a quantitative change in Meissner corpuscles, a change of the mechanical properties of the dermis, degenerative transformations of the Pacinian corpuscles, demyelination and fibre loss in peripheral nerves, and degenerative changes in the central nervous system.[10] The lower dynamic threshold values obtained probably result from measuring the quickly adapting fibre system instead of the slowly adapting fibres as in static assessment.[11, 12] However, in our patient group, no method seemed to be superior: the ability to detect a patient was

equally low in static and dynamic assessments. A moderate correlation between the two-point discrimination values and the arm grade of the ODSS was found in the patients who were mildly affected by a polyneuropathy. Apparently, the aesthesiometer, measuring an impairment quality, translates its findings to daily activity deficits. Also, the aesthesiometer shows its convergent validity through correlation with the WEST.

Despite the findings in this study, some methodological issues should be addressed. Firstly, although most polyneuropathies are known to be length dependent, we have chosen to examine two-point discrimination at the index finger. In a pilot study, we found great variability in threshold values with low inter-rater scores when tested at the hallux, whereas a good reliability at the index finger has been reported.[13] Secondly, despite its simplicity, its validity, and its demonstrated reliability as part of the ISS, two-point discrimination assessment needs further clinimetric evaluation.[1] Currently, the reliability and its responsiveness are under investigation in patients with inflammatory polyneuropathies (***P*eri*N*om*S*** study).

Owing to its low sensitivity in mildly affected patients, we suggest using two-point discrimination assessment as part of a more extensive sensory examination, using its revised reference values now available.

References

1. Merkies, I.S., et al., *Psychometric evaluation of a new sensory scale in immune-mediated polyneuropathies. Inflammatory Neuropathy Cause and Treatment (INCAT) Group.* Neurology, 2000. **54**(4): p. 943-9.
2. Louis, D.S., et al., *Evaluation of normal values for stationary and moving two-point discrimination in the hand.* J Hand Surg [Am], 1984. **9**(4): p. 552-5.
3. Shimokata, H. and F. Kuzuya, *Two-point discrimination test of the skin as an index of sensory aging.* Gerontology, 1995. **41**(5): p. 267-72.
4. Kaneko, A., N. Asai, and T. Kanda, *The influence of age on pressure perception of static and moving two-point discrimination in normal subjects.* J Hand Ther, 2005. **18**(4): p. 421-4, quiz 425.
5. Desrosiers, J., et al., *Hand sensibility of healthy older people.* J Am Geriatr Soc, 1996. **44**(8): p. 974-8.
6. Hermann, R.P., C.B. Novak, and S.E. Mackinnon, *Establishing normal values of moving two-point discrimination in children and adolescents.* Dev Med Child Neurol, 1996. **38**(3): p. 255-61.
7. Martina, I.S., et al., *Measuring vibration threshold with a graduated tuning fork in normal aging and in patients with polyneuropathy. European Inflammatory Neuropathy Cause and Treatment (INCAT) group.* J Neurol Neurosurg Psychiatry, 1998. **65**(5): p. 743-7.
8. Weinstein, S., *Fifty years of somatosensory research: from the Semmes-Weinstein monofilaments to the Weinstein Enhanced Sensory Test.* J Hand Ther, 1993. **6**(1): p. 11-22; discussion 50.
9. Merkies, I.S., et al., *Clinimetric evaluation of a new overall disability scale in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2002. **72**(5): p. 596-601.
10. Wickremaratchi, M.M. and J.G. Llewelyn, *Effects of ageing on touch.* Postgrad Med J, 2006. **82**(967): p. 301-4.
11. Dellon, A.L., *The moving two-point discrimination test: clinical evaluation of the quickly adapting fiber/receptor system.* J Hand Surg [Am], 1978. **3**(5): p. 474-81.
12. Menier, C., R. Forget, and J. Lambert, *Evaluation of two-point discrimination in children: reliability, effects of passive displacement and voluntary movements.* Dev Med Child Neurol, 1996. **38**(6): p. 523-37.
13. Dellon, A.L., S.E. Mackinnon, and P.M. Crosby, *Reliability of two-point discrimination measurements.* J Hand Surg [Am], 1987. **12**(5 Pt 1): p. 693-6.

Chapter 2.2

Revised normative values for grip strength with the Jamar dynamometer

Martine J.H. Peters[1], Sonja I. van Nes[2], Els K. Vanhoutte[1], Mayienne Bakkers[1], Pieter A. van Doorn [2], Ingemar S.J. Merkies[1,2,3],and Catharina G. Faber[1], on behalf of the PeriNomS Study group

[1]Department of Neurology, Maastricht University Medical Centre, Maastricht, the Netherlands,
[2]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands,
[3]Department of Neurology, Spaarne hospital, Hoofddorp, the Netherlands

J Peripher Nerv Syst. 2011 Mar; 16(1): 47-50

Abstract

The Jamar dynamometer has been widely used in various chronic illnesses and has demonstrated its strength as a potential prognostic indicator. Various stratified normative values have been published using different methodologies, leading to conflicting results. No study used statistical techniques considering the non-Gaussian distribution of the obtained grip strength values. Jamar grip strength was assessed in 720 healthy participants, subdivided into seven age decade groups consisting of at least 50 men and 50 women each. Normative values (median and 5th values) were calculated using quantile regressions with restricted cubic spline functions on age. Possible confounding personal factors (hand dominance, height, weight, hobby and job categorisation) were examined. Clinically applicable revised normative values for the Jamar dynamometer, stratified for age and gender, are presented. Hand dominance had no influence. Other personal factors only minimally influenced final values. This study provides revised normative grip strength values for the Jamar dynamometer.

Introduction

Distal weakness generally predominates in patients with polyneuropathy and may contribute to deficits in daily activities and social participation.[1, 2] Therefore, reliable assessment of grip strength is very important to capture clinical changes in interventional studies in these disorders. Assessment of grip strength has also demonstrated to be a prognostic indicator of clinical and functional recovery in other diseases affecting hand function.[3, 4] Two devices are in widespread use, the Vigorimeter and the Jamar dynamometer. Normative values for the Vigorimeter have been published using new statistical techniques.[2] Various stratified normative values for the Jamar dynamometer have been published; however, using different methodologies leading to conflicting results. No information has been provided regarding the statistical analyses for the first report on normative values for the Jamar.[5] In a descriptive meta-analysis, means and 95% confidence intervals of normative values per age category were listed.[6] Unfortunately, the methodology among the 12 recruited studies differed substantially, making the proposed normative values not suitable. In a recent study, the reported normative values were based on means (+SD), while the normative values seemed quite skewed showing a curvilinear relationship to age.[7] On the basis of these observations, we examined a large cohort of healthy individuals from 20 to 96 years of age to obtain revised normative values for the Jamar dynamometer using statistical techniques that capture the non-Gaussian distribution of the obtained grip strength values.

Patients and methods

A total of 720 healthy participants (20-96 years) were recruited from university, hospital and secondary school personnel, homes for the elderly and sports clubs. They volunteered on locally spread pamphlets requesting participation and were stratified for age and gender forming seven age decade groups (20-29, 30-39,..., ≥80 years), each consisting of 50 men and 50 women. Inclusion criteria were: independence in daily living, absence of any impairment affecting upper limb function (e.g., joint deformities), absence of impaired sensory function (e.g., burning, tingling, numbness of hands or feet), no concomitant disease or medication possibly causing polyneuropathy (e.g., diabetes mellitus, chemotherapy). All participants were neurologically examined with emphasis on upper extremity function and signs of peripheral nervous system deficits (e.g., muscle strength (Medical Research Council grading), tendon reflexes, light touch, pin-prick, vibration sense (Rydel-Seiffer tuning fork)).[8] Eligibility was obtained if examination was normal.

Assessment tools

The portable Jamar hydraulic hand dynamometer (Sammons Preston, Rolyan, Bolingbrook, IL, USA) was used, with the second handle position for all subjects. The presented scores are expressed in pounds force.

Test procedures

The local medical ethics committee approved the protocol. All participants gave written informed consent. Through interview we collected anthropometric data (body height, weight) and hand dominance. Participants were requested to categorise their profession and hobby as being physically light (coded: 0), moderate (1) or heavy (2), if applicable.

The Jamar dynamometer was placed in the right or left hand randomly and was held loosely around the readout dial by the examiner to prevent dropping.[5] In alternating order, three maximum voluntary grip strength contractions were taken for each hand. The mean value of each hand was used for analysis.[9] All participants were examined in a standardised position.[10]

Statistics

After stratification for age and gender, revised normative values (median and 0,05 quantile values, corresponding to a specificity of 95%) for the Jamar dynamometer were calculated using quantile regression analyses with restricted cubic spline functions on age.[11] In each gender, multivariate quantile regression analyses were performed, with height, weight, categorisation of profession and hobby as the independent variables on the calculated 5th quantile normative values (dependent variable).[2, 11, 12] All analyses were performed using STATA 11.0 for Windows.

Results

Descriptive data of all 720 participants are presented in table 1. Men were generally stronger than women. There was a significant curvilinear age-dependent decrease of normative values in both genders. Maximum median grip strength was reached among women aged 30-39 and among men aged 40-49. Furthermore, no significant difference in grip strength between the dominant and non-dominant hand was found (overall median difference of 1 pound). Therefore we developed one normative values graph for both hands for each gender separately (table 2; figure 1).

Multivariate regression analysis did not show any significant impact of height, weight, profession or hobby categorisation on the 5th percentile cut-off normative values in women. In men, only hobby categorisation was related to the 5th percentile cut-off normative values; however, explaining only a small portion of 7% (p=0.008).

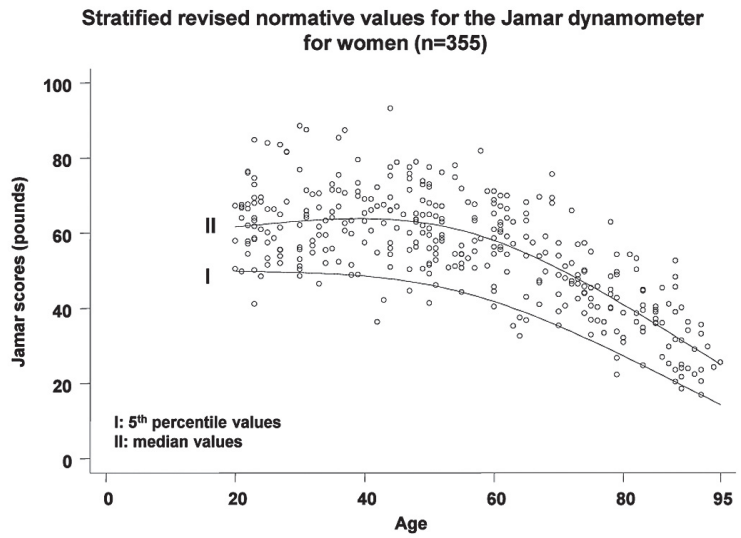
Table 1. Descriptive data of healthy participants

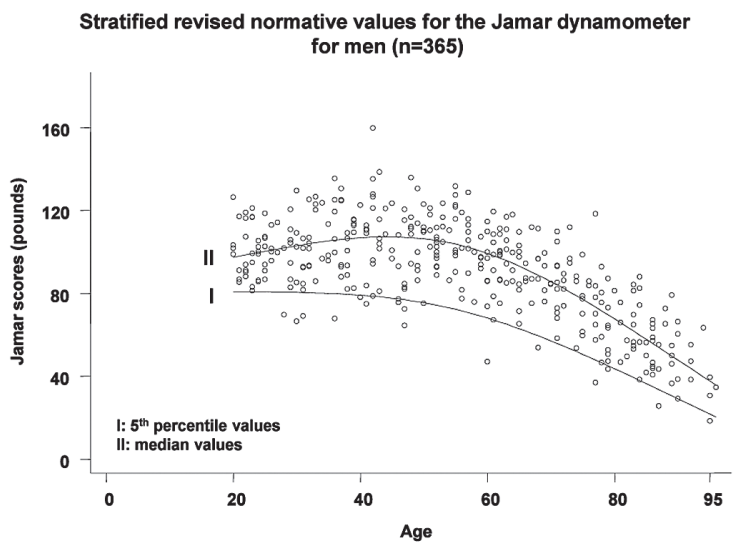
	Women	Men	Total
Participants, n	355	365	720
Age Mean (SD) range (year)	 54.7 (21.1) 20-95	 55.1 (20.5) 20-96	 54.9 (20.8) 20-96
Hand dominance			
Right (%)	90	87	88
Left (%)	7	10	8
Ambidextrous (%)	3	4	3
Height			
Mean, in cm (SD)	166 (7.1)	179 (8.3)	173 (10.1)
Range (cm)	140-184	150-202	140-202
Weight			
Mean, in kg (SD)	68.2 (12.4)	82.1 (12.9)	75.2 (14.4)
Range (kg)	43-123	45-135	43-135
Profession categorisation			
Light, n (%)	115 (32.4)	136 (37.3)	251 (34.9)
Moderate, n (%)	95 (26.8)	83 (22.7)	178 (24.7)
Heavy, n (%)	25 (7.0)	23 (6.3)	48 (6.7)
Not applicable, n (%)	120 (33.8)	123 (33.7)	243 (33.8)
Hobby categorisation			
Light, n (%)	197 (55.5%)	163 (44.7%)	360 (50.0%)
Moderate, n (%)	143 (40.3%)	171 (46.9%)	314 (43.6%)
Heavy, n (%)	15 (4.2%)	31 (8.5%)	46 (6.4%)

Table 2. Normative grip strength values for the Jamar dynamometer for clinical use (in pounds)

	Women (n=355)			Men (n=365)		
Age (years)	Number of subjects	5th percentile Jamar values (pounds)	Median Jamar values (pounds)	Number of subjects	5th percentile Jamar values (pounds)	Median Jamar values (pounds)
20-29	51	50	62	50	81	100
30-39	50	49	64	51	80	105
40-49	50	48	63	50	78	107
50-59	51	45	61	54	73	104
60-69	49	40	56	58	64	95
70-79	50	32	46	50	51	77
≥ 80	54	22	34	52	34	54

Figure 1. Revised grip strength values with the Jamar dynamometer for the dominant and non-dominant hands combined in healthy women (n= 355, 90.1% right-handed) and healthy men (n= 365, 86.6% right-handed).





Legend to figure 1. The upper lines in each graph correspond to the calculated median grip strength values. The lower lines in each graph represent the 0,05 quantile reference values. These values were obtained in each gender separately using quantile regressions with restricted cubic spline functions. Note: the y-axis for men and women differ

Discussion

This study provides revised normative values for the Jamar dynamometer in adults. We used robust statistical techniques to present the correlation between age and grip strength values.[11] Normative values for the Jamar dynamometer have been provided before.[5-7, 13, 14] Unfortunately, all these values were obtained by calculating means of sequential age portions, without taking into account the non-Gaussian distribution of the obtained grip strength values, as recommended for these kinds of continuous data.[2, 11, 12] Grip strength showed a curvilinear relationship to age.[5, 7, 14] In both genders, a marked decline of grip strength was seen after 60 years of age, which is in conformity with earlier reports.[5, 7] Our data showed hardly any difference in grip strength between the dominant and non-dominant hand, which is in contrast with earlier papers reporting up to 33% difference.[5, 7] A difference in statistical approach could be an explanation for the obtained differences. At a workshop on outcome measures in immune-mediated neuropathies, it was argued whether the Vigorimeter, with its soft squeezing bulb and particularly used in Europe, would have a preference over the Jamar dynamometer, mainly used in the United States.[15] At current stage, there is no consensus regarding which instrument should be used. Therefore, a comparison study of these two instruments is warranted, focusing primarily on their responsiveness findings and on patients' preference. These efforts are currently being undertaken as part of an international, multi-centre collaborative study (*P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study).

Our calculated 5th percentile normative cut-off grip strength data did not show any correlation with parameters like weight, height, occupation, except for hobby categorisation in men, which is in contrast with recent literature.[7, 16] However, the studies by Werle *et al.* and Angst *et al.* have used a different approach: they have calculated a correlation between the obtained raw grip strength scores and other parameters (height, weight, occupation) and have not used a cut-off score for these normative values, which may explain the differences with our paper.[7, 16] In conclusion, the current study provides robust clinically applicable normative 5th percentile grip strength values for the Jamar dynamometer.

References

1. Kleyweg, R.P., F.G. van der Meche, and P.I. Schmitz, *Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barré syndrome.* Muscle Nerve, 1991. **14**(11): p. 1103-9.
2. Merkies, I.S., et al., *Assessing grip strength in healthy individuals and patients with immune-mediated polyneuropathies.* Muscle Nerve, 2000. **23**(9): p. 1393-401.
3. Langhammer, B., B. Lindmark, and J.K. Stanghelle, *Stroke patients and long-term training: is it worthwhile? A randomized comparison of two different training strategies after rehabilitation.* Clin Rehabil, 2007. **21**(6): p. 495-510.
4. Formsma, S.A., C.K. van der Sluis, and P.U. Dijkstra, *Effectiveness of a MP-blocking splint and therapy in rheumatoid arthritis: a descriptive pilot study.* J Hand Ther, 2008. **21**(4): p. 347-53.
5. Mathiowetz, V., et al., *Reliability and validity of grip and pinch strength evaluations.* J Hand Surg Am, 1984. **9**(2): p. 222-6.
6. Bohannon, R.W., et al., *Reference values for adult grip strength measured with a Jamar dynamometer: a descriptive meta-analysis.* Physiotherapy, 2006. **92**: p. 11-15.
7. Werle, S., et al., *Age- and gender-specific normative data of grip and pinch strength in a healthy adult Swiss population.* J Hand Surg Eur Vol, 2009. **34**(1): p. 76-84.
8. Martina, I.S., et al., *Measuring vibration threshold with a graduated tuning fork in normal aging and in patients with polyneuropathy. European Inflammatory Neuropathy Cause and Treatment (INCAT) group.* J Neurol Neurosurg Psychiatry, 1998. **65**(5): p. 743-7.
9. Mathiowetz, V., *Grip and pinch strength measurements*, in *Muscle strength testing. Instrumented and non-instrumented systems*, L.R. Amundsen, Editor. 1990, Churchill Livingstone: New York. p. 163-177.
10. American Society of Hand Therapists, *Clinical assessment recommendations*. 2nd ed. 1992, Chicago.
11. Gould, W.W. and W.H. Rogers, *Quantile regression as an alternative to robust regression. Proceedings of the statistical computing section.* 1994, Alexandria; Virginia: American Statistical Association
12. Herndon J.E. and F.E. Harrel Jr, *The restricted cubic spline hazard model.* Comm Stat Theory Meth, 1990. **19**: p. 639-663.
13. Harkonen, R., M. Piirtomaa, and H. Alaranta, *Grip strength and hand position of the dynamometer in 204 Finnish adults.* J Hand Surg Br, 1993. **18**(1): p. 129-32.
14. Massy-Westropp, N., et al., *Measuring grip strength in normal adults: reference ranges and a comparison of electronic and hydraulic instruments.* J Hand Surg Am, 2004. **29**(3): p. 514-9.
15. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
16. Angst, F., et al., *Prediction of grip and key pinch strength in 978 healthy subjects.* BMC Musculoskelet Disord, 2010. **11**: p. 94.

Chapter 3

Rasch-built outcome measures

Chapter 3.1

Explaining Rasch to neurologists

Sonja I. van Nes[1], Els K. Vanhoutte[2], Pieter A. van Doorn[1], Mieke Hermans[2], Mayienne Bakkers[2], Krista Kuitwaard[1], Catharina G. Faber[2], Ingemar S.J. Merkies[1,2,3], on behalf of the PeriNomS study group

[1]Department of neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of neurology, Maastricht University Medical Centre, Maastricht, the Netherlands
[3]Department of neurology, Spaarne Hospital, Hoofddorp, the Netherlands

Adapted from supplemental material - Neurology 2011;76:337-345

Introduction

This brief introduction is intended to help clinicians and scientists in neurology understand the underlying principles of the Rasch method. Immune-mediated neuropathies will be used as an example, however, this writing is essentially applicable to all (neurological) illnesses. A review was published in 2008 highlighting the use of all types of outcome measures in published randomised controlled trials that included patients with various forms of immune-mediated neuropathies (see also chapter 1.2).[1] This paper (chapter 1.2) also presented the basic clinimetric requirements (like validity, reliability, and responsiveness) for any outcome measure to be selected for use in a trial.[1, 2] Since then, additional randomised trials have been published demonstrating the efficacy of medical interventions in patients with chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) and monoclonal gammopathy of undetermined significance related neuropathies (MGUSP) using similar ordinal based outcome measures.[3-8] All outcome measures used to date in immune-mediated neuropathy medical interventional trials are based on the classical test theory (CTT).[9] DeVellis has summarised various aspects of CTT including its disadvantages: outcome measures based on CTT may constitute items that are arbitrarily collected with response options generally based on ordinal Likert-type choices.[9] As an example, physicians often consider a one point response change for an item from 0 to 1 equivalent to a one point change from 2 to 3. However, because the response options are ordinal based, the true distance between the response categories is not known and most probably unequal (figure 1).

Figure 1. Example of an outcome measure based on classical test theory with four ordinal Likert response options per item

	impossible to perform 0	with great difficulty 1	with slight difficulty 2	Easily performed 3
washing face	◄	► ?	◄	► ?
stand up from a chair				
walk 1 flight of stairs				
take a shower				

Legend to figure 1. A one point change from 0 to 1 for an item is considered being equivalent to a one point change from 2 to 3; however, because the response options are ordinal based, the equivalence is highly unlikely.

Also, patients are requested to complete all items, even though some may be irrelevant or inappropriate for their level of ability. Often, scores of the items of a scale are summed and the obtained data generally treated as if they were linear; frequently being exposed to parametric analyses. Creating a sum of the item scores

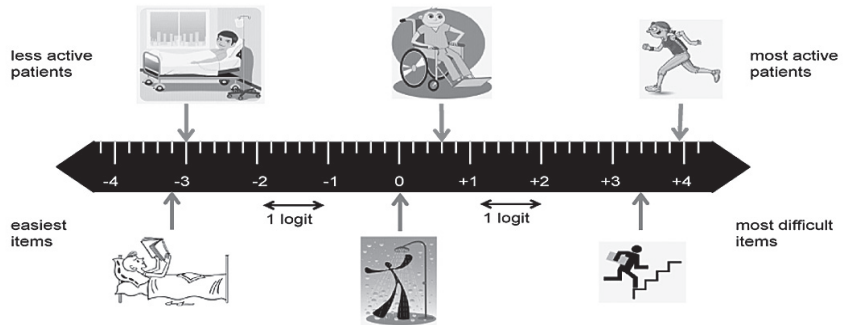
also assumes equal relevance ('weight') of each item, which is highly unlikely.[10] Based on these shortcomings, CTT-based outcome measures may limit the comparison of patients and study results. Ordinal scales lack linearity.

Rasch as a modern clinimetric vehicle

Considering the shortcomings of the CTT, it is clear that a modern scientific approach is needed for the evaluation and construction of outcome measures to improve the findings in interventional trials. Using interval measures with a linear construct instead of ordinal scores would give a true reflection of disease impact, of differences between individuals and groups, and treatment effects. One of the widely used approaches is the Rasch method, which was introduced by the Danish mathematician Georg Rasch.[11, 12]

The Rasch method is based on logic assumptions. For example, it is very logical to assume that walking a flight of stairs will be a much more difficult task to accomplish compared to washing one's face. It is also logical to assume that a Guillain-Barré syndrome (GBS) patient that is bed-bound due to severe weakness will have more problems performing daily activities than a GBS patient walking around with only an ankle orthosis. From a more statistical background, the Rasch model states that the probability of a patient being able to 'correctly answer or complete an item or task' is a logistic function of the difficulty of the task and the ability of the patient to accomplish it.[11] Rasch analysis transforms obtained ordinal scores into interval measures and places both items' and patients' parameter estimates on the same log-odds unit (logit) scale (figure 2). Therefore a less affected patient (a patient with a higher ability) will have a greater chance to complete a more difficult item when compared to a patient that is more disabled.

Figure 2. Example of a Rasch-built interval activity outcome measure with linearly weighted estimates of patient ability and item difficulty



Legend to figure 2. The Rasch model compares the item response patterns of individuals to the entire sample of patients being examined to estimate person ability and item difficulty and places both item and person estimates on the same logit scale.

This chapter aims to help clinicians and scientists in neurology increase their understanding of the various steps of the Rasch methodology.

Ordering item difficulty estimates and person ability estimates on the same ruler

The statistical calculations and procedures of ordering the difficulty of the items and the ability of patients on the same ruler is based on the so-called Guttman scaling.[13] As illustrated in figure 3A, the Guttman scaling codes items as 0 ('no/ not able') or 1 ('yes/ able'). Figures 3B and 3C demonstrate the ordering of the difficulty of items based on Guttman scaling from most easy (coded 01111- sumscore 4) to most difficult (00001- sumscore 1). Persons are also ordered according to their ability level: patient E having a higher sumscore thus a greater ability than patient A (figure 3D and 3E).

Figure 3A. Guttman scaling

	Item 1	Item 2	Item 3	Item 4	*Sumscore*
Person A	0	0	0	0	0
Person B	1	0	0	0	1
Person C	1	1	0	0	2
Person D	1	1	1	0	3
Person E	1	1	1	1	4
Sumscore	4	3	2	1	

Legend to figure 3A. As an example, item 1 refers to the question 'are you able to wash your face'. Person A responded 'No, I am not able to wash my face' coded as 0. Person B responded 'yes, I am able to wash my face' coded as 1.

Figure 3B. Easiest item based on Guttman scaling

item difficulty

	Item 1	Item 2	Item 3	Item 4	*Sumscore*
Person A	**0**	0	0	0	0
Person B	**1**	0	0	0	1
Person C	**1**	1	0	0	2
Person D	**1**	1	1	0	3
Person E	**1**	1	1	1	4
Sumscore	**4**	3	2	1	

Easiest item

Legend to figure 3B. Person A responded 0 ('no, I am not able to…') to item 1. However, person B-E responded 1 ('yes, I am able to…') to item 1. Therefore, for this population, item 1 is relatively easy.

Figure 3C. Most difficult item based on Guttman scaling

item difficulty

	Item 1	Item 2	Item 3	**Item 4**	*Sumscore*
Person A	0	0	0	**0**	0
Person B	1	0	0	**0**	1
Person C	1	1	0	**0**	2
Person D	1	1	1	**0**	3
Person E	1	1	1	**1**	4
Sumscore	4	3	2	**1**	

Most difficult item

Legend to figure 3C. Person A-D responded 0 ('no, I am not able to...') to item 4. Only person E responded 1 ('yes, I am able to...') to item 4. Therefore, for this population, item 4 is rather difficult.

Figure 3D. Guttman scaling reveals the most able patient completing all items

patient's ability

	Item 1	Item 2	Item 3	Item 4	***Sumscore***
Person A	0	0	0	0	0
Person B	1	0	0	0	1
Person C	1	1	0	0	2
Person D	1	1	1	0	3
Person E	**1**	**1**	**1**	**1**	**4**
Sumscore	4	3	2	1	

Most able

Legend to figure 3D. Person E responded 1 ('yes, I am able to...') to all items corresponding to being able to perform all activities captured by item 1-4. Therefore, person E has a high ability level compared to other persons in this population.

Figure 3E. Guttman scaling reveals the least able patient completing none of the items

patient's ability

	Item 1	Item 2	Item 3	Item 4	***Sumscore***
Person A	**0**	**0**	**0**	**0**	**0**
Person B	1	0	0	0	1
Person C	1	1	0	0	2
Person D	1	1	1	0	3
Person E	1	1	1	1	4
Sumscore	4	3	2	1	

Least able

Legend to figure 3E. Person A responded 0 ('no, I am not able to...') to all items corresponding to not being able to perform any of the activities captured by item 1-4. Therefore, person A has a low ability level compared to other persons in the population examined.

Figure 4A-J illustrates in a schematic way the statistical steps taken by the Rasch model to order items and patients on one ruler. Suppose we are examining n=100 patients with an immune-mediated neuropathy using a daily activity scale that includes 4 items (A, B, C, and D) with ordinal response options ranging from 0 ('impossible to perform'), to 1 ('very difficult to perform'), to 2 ('difficult to perform'), and to 3 ('easy to perform') (figure 4A).

Figure 4A.

	Impossible 0	Very difficult 1	Difficult 2	Easy 3
Item A				
Item B				
Item C				
Item D				

■ **N=100** patients examined

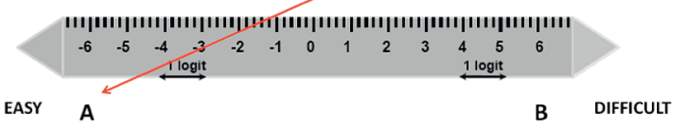
Figure 4B. Number of patients per response option after completion of the 4-item daily activity scale by 100 patients

	Impossible 0	Very difficult 1	Difficult 2	Easy 3
Item A		**1**	**4**	**95**
Item B	98	2		
Item C	10	40	36	14
Item D	50	35	10	5

■ **N=100** patients examined

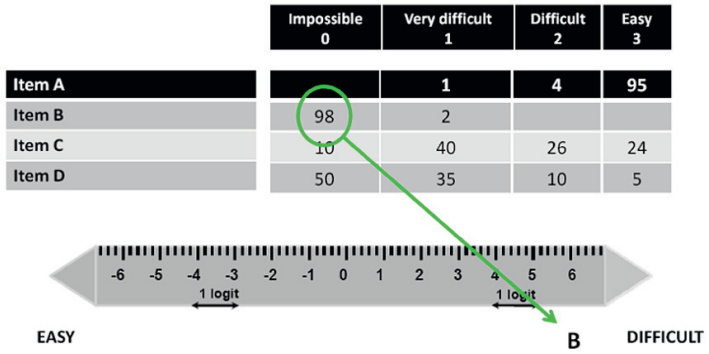
Figure 4C. Ordering items according to their difficulty level: location of the easiest item (item A) on the ruler

	Impossible 0	Very difficult 1	Difficult 2	Easy 3
Item A		**1**	**4**	**95**
Item B	98	2		
Item C	10	40	26	24
Item D	50	35	10	5



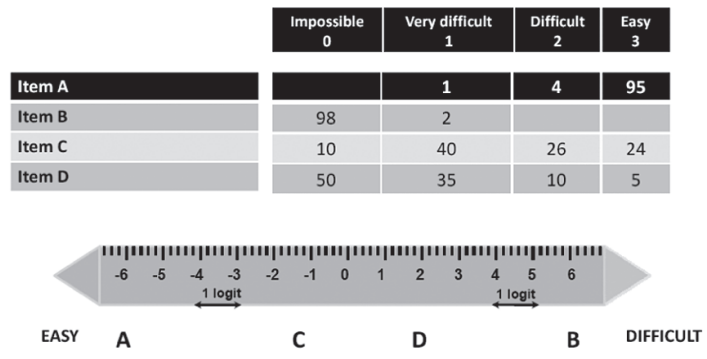
As can be seen, item A turns out to be the 'easiest item' based on the highest percentage (95%) of patients examined choosing response option 3 ('easy to perform') for this item (figure 4C). Conversely, 98 of the 100 patients scored 0 ('impossible') on item B, making this item the most difficult to accomplish (figure 4D).

Figure 4D. Ordering items according to their difficulty level: location of the most difficult item (item D) on the same ruler



Subsequently, figure 4E show a stepwise ordering of all items by the Rasch model on the same ruler.

Figure 4E. Ordering the items A, B, C, and D on one ruler according to their difficulty level

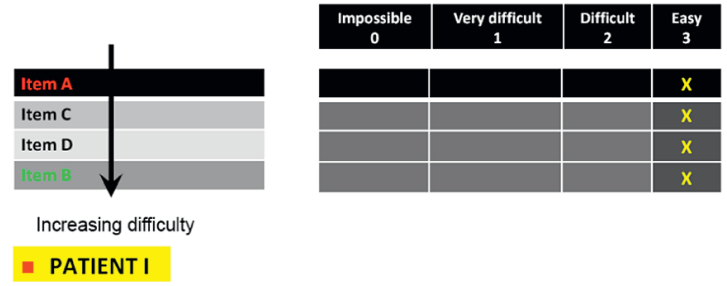


Legend to figure 4E. Since 50% of the patients had a score of 0 ('impossible to perform') on item D and only 10% on item C, it is obvious that item D is more difficult to accomplish compared to item C. The Rasch model estimates their location ('weight') as expressed in logits.

Figures 4F through 4J illustrate the stepwise ordering of patients on the same ruler. To maintain clarity, we will be comparing the results of 3 patients completing the

above mentioned daily activities scale as an example. Figure 4F shows the same scale, but now the items are listed based on their weights ('difficulty to accomplish'). As shown, patient I scored 3 ('easy to perform') on all four items.

Figure 4F. Ordering patients according to their ability level, starting with patient I (yellow)



The scores for patients II and III are presented in figures 4G and 4H.

Figure 4G. Patient II (green)

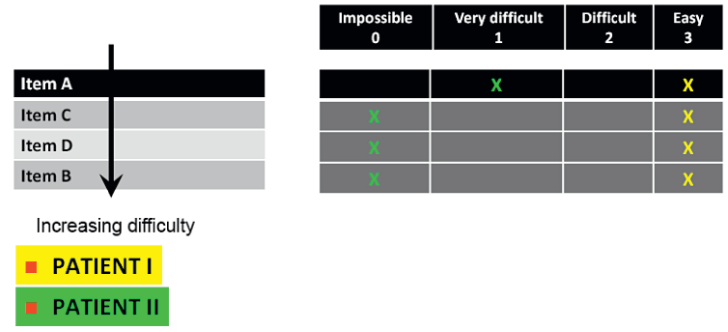
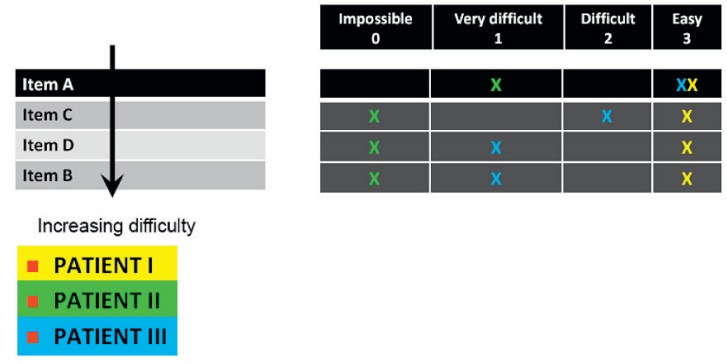


Figure 4H. Patient III (blue)



Based on the obtained results in these three patients it is concluded that patient I (yellow) has the highest ability, scoring the maximum on all daily activities scale's items. On the other hand, patient II (green) had the lowest scores when compared with the other patients, hence demonstrating having the lowest ability of the three patients. Their ordering on the same metric is illustrated in the figures 4I and 4J, thus obtaining a final model with all items and patients equated on the same metric (see also figure 2).

Figure 4I. Starting with the patient with the highest ability: patient I (yellow), because all scores of this patient were 3 ('easy to perform')

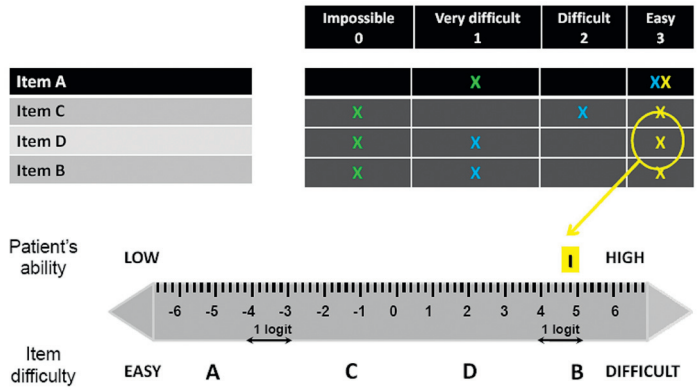
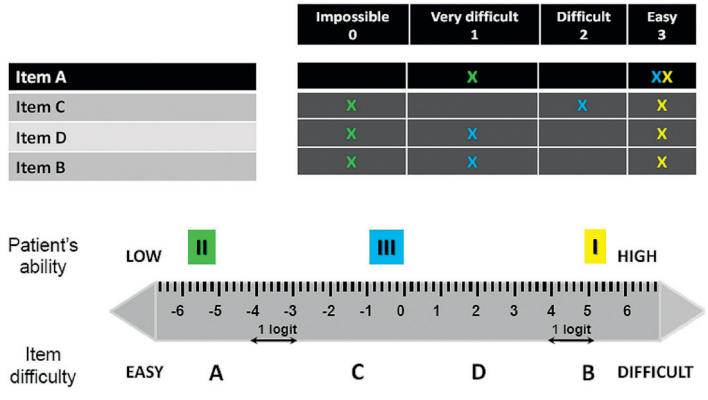


Figure 4J. Final schematic ordering of items'difficulties and patients'abilities on the same metric by the Rasch model



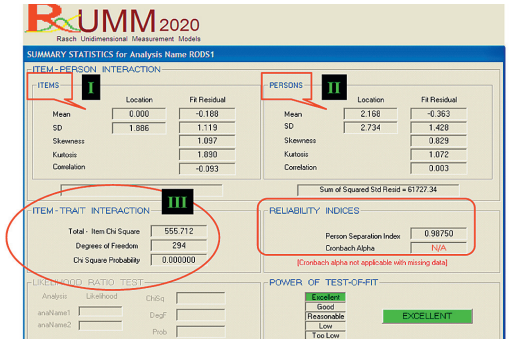
Rasch model expectations

To obtain an interval scale with a linear construct based on the Rasch method, all items and persons as part of the model need to fulfil several model expectations.[11,12] These expectations could be seen as *check-points* in the process of creating a scale. Only when all expectations are met, an ideal interval scale can be created. The Rasch Unidimensional Measurement Models (RUMM), used for this purpose, elegantly integrates the various measurement issues and provides the researcher with numerous explanatory graphs.[14] The following applications of RUMM are used to examine whether the data fit the model expectations. Items or patients not fulfilling these requirements should be removed or subjected to re-adjustments to fit the model. The statistical background of the RUMM will only be addressed briefly as part of a control panel of the software, which automatically provides the results of the analyses performed (figure 5).

Fit statistics

The RUMM has three overall fit statistics that have to be fulfilled. Two are item-person interaction statistics transformed to approximate a z-score. Therefore, if the items and persons fit the model, a mean around zero and a standard deviation of 1 would be expected (figure 5, sections I and II).[12] A third fit statistic score is an item-trait interaction statistic reported as a chi-square, reflecting the property of invariance across the trait to be measured. A non-significant chi-square indicates that the hierarchical ordering of the items does not vary across the trait, thus fulfilling the required property of invariance (figure 5; section III).

Figure 5. Example of summary statistics as provided by RUMM during model creation



Fit residuals

Individual person- and item-fit statistics are also examined both as residuals (as a summation of individual patient and item deviations from expected model scores) and as chi-square statistics (see figure 5, sections II and III). In the former case,

residuals between ±2.5 are considered adequate fit to the model. In the latter case a chi-square statistic is available for each item. Summation of the overall chi-square for items gives the item-trait interaction statistic.

Internal reliability studies

Internal consistency of the scale is determined with the person separation index (PSI) or Cronbach's alpha (the latter is only possible when there are no missing values) using the logit scores for each person. A value of ≥ 0.7 is considered consistent with the scale being able to differentiate between at least 2 groups of patients and is seen as the minimum requirement for measurement.[15]

Sample size calculations

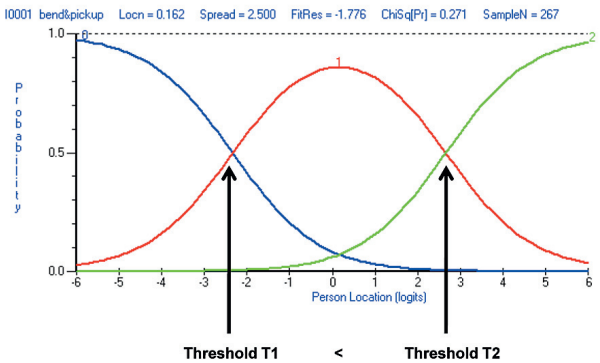
Professor Linacre, one of the well-known researchers in the field of Rasch analyses, has extensively examined the ideal sample size needed to have a stable model through modelling. From his publications, a sample size of approximately 250 is needed to obtain a 99% confidence with a stable item calibration within ± 0.5 logits, hence providing a stable model.[16]

Additional requirements for model fitting

Threshold examination

The term threshold refers to the point between two adjacent response categories where either response is equally probable. That is the point where, for example in case of the Rasch-built overall disability scale (R-ODS), the probability of scoring 0 ('impossible to perform') and 1 ('performed with difficulty') is 50/50 for an item (figure 6). In this example the individual patient is able to differentiate between the response options 0, 1, and 2, since the two threshold locations are ordered (figure 6).

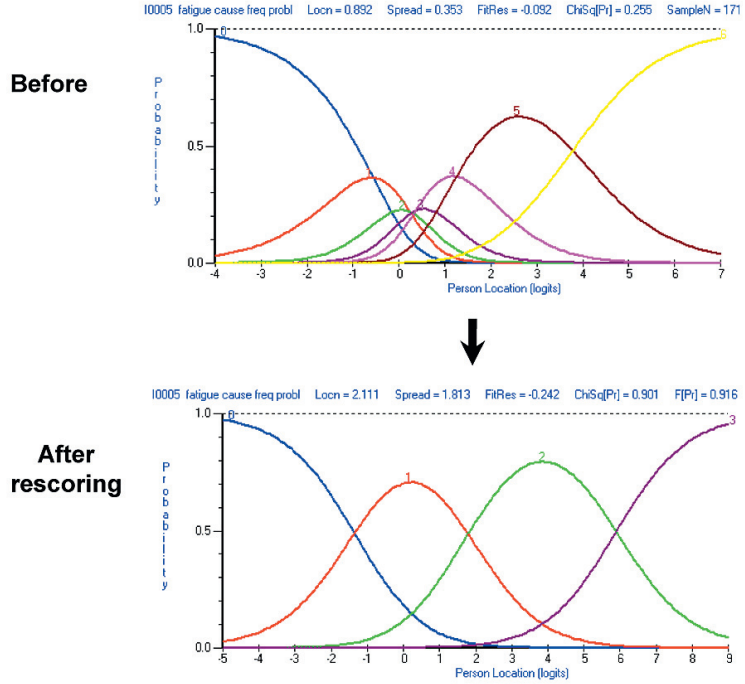
Figure 6. Category probability curves demonstrating the ideal ability of patients to discriminate between response options for the item 'bend forward and pickup an object'



Legend to figure 6. The Rasch model translates the response options graphically for each item of a scale; in this example, the response options are defined as: 0 ('impossible to perform') blue line, 1('performed with difficulty') - red line, and 2 ('easily performed') - green line.
Note: Threshold T1 < T2, indicating good ordering of these by the patients.[12]

One of the most common sources of item misfit concerns respondents inconsistent use of response options. This results in what is known as 'reversed or disordered thresholds'. Disordered thresholds occur when respondents have consistently difficulty discriminating between response options. This can occur when there are too many response options, or when the labelling of options is potentially confusing or open to misinterpretation (e.g., great - some - little). An example of disordered thresholds is given in figure 7, demonstrating the inability of patients to properly discriminate between the response options for the item 'fatigue causes frequent problems for me' as part of the fatigue severity scale (see also chapter 3.2).[17] The response options range from: 0 ('strongly disagree'), 1 ('mainly disagree'), 2 ('partially disagree'), 3 ('do not agree/disagree'), 4 ('partially agree'), 5 ('mainly agree'), and 6 ('strongly agree'). In particular, the response categories 1 to 4 were clustered, thus showing a disordered threshold pattern. In order to improve model fit, we subsequently collapsed the response options to obtain ordered thresholds (figure 7; see also the pattern presented in figure 6). In the current example, the model suggested rescoring the categories 1 through 5 into 2 response options (changing the total response options from 0/1/2/3/4/5/6 to 0/1/2/3).

Figure 7. Example of 'disordered threshold' at initial examination and 'ordered thresholds' after rescoring of the response categories.



Local dependency

Residual correlations between items within the same scale are a source of misfit. Local dependency reflected by residual correlations arises when items are linked such that the response on one item may be dependent upon the response to another. This finding inflates reliability and the final scale score in a particular direction.[12] Figure 8 illustrates local dependency by looking at parts of the motor question of the SF-36 health survey.[18] In cases of local dependency, the researcher may consider removing items or creating a subset of correlating items to improve model fit.

Figure 8. Local dependency demonstrated through motor question as part of the SF-36 health survey

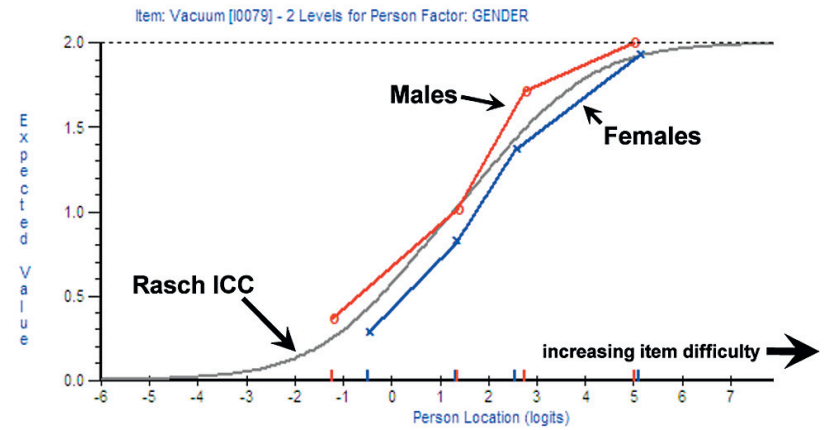
Does your health now limit you in these activities? If so, how much?	Yes, limited a lot	Yes, limited a little	No, not limited at all
Walking more than a mile?	1	2	3
Walking several hundred yards?	1	2	3
Walking one hundred yards?	1	2	3

Legend to figure 8. If a patient scores 3 ('no, not limited at all') on the question 'does your health limit you now in walking more than a mile?', this automatically has a bearing upon the answer on the other two question; in other words: this patient will also be scoring 3 on the questions 'does your health limit you now in walking several hundred yards?' and 'does your health limit you now in walking one hundred yards?'. Since the results of these 3 questions are linked, the obtained scores will inflate reliability and the final scale' score in a particular direction.

Item bias

A scale should always work in the same way irrespective of which personal factor (e.g., gender) is being examined.[19] For example, if men and women with equal ability levels (equal limitations due to their immune-mediated neuropathy) respond systematically differently to a daily activity item, then this item would be considered to display differential item functioning (DIF) = item bias, and would violate the requirement for unidimensionality.[12] Figure 9 illustrates DIF on personal factor 'gender' for the item 'vacuum cleaning'. As can be seen, females (blue line) experienced this activity as more difficult to accomplish than males (red line). Subsequently, this item was removed from the model. The researcher may consider to omit such an item or to split this into two (a vacuum cleaning item specifically for males and females, separately).

Figure 9. Item characteristic curve for item 'vacuum cleaning' related to personal factor 'gender'



Legend to figure 9. ICC = item characteristic curve. This picture shows how DIF puts the males (red line) to the left (easier to perform) side of the Rasch ICC curve, and the females (blue line) to the right (more difficult to perform) side.

Calibration of the scale – potential drawbacks

Items and persons are selected or discarded based on the above mentioned stepwise approach. As a result this determines the final locations of items and persons on the ruler (calibration of the scale). Therefore, the study population should represent the patients you would like to evaluate in the future. Items should be unambiguously constructed to prevent misinterpretation. A questionnaire should contain written instructions. Nevertheless, there might be some bias of the responses given. Patients may not disclose a proper response without being interviewed personally. However, inconsistent responses to certain items will probably result in the item being omitted due to misfit statistics. Once data fit Rasch model expectations, logits of person estimates can be used as an interval level variable in parametric statistics.

In conclusion, modern clinimetric methods such as Rasch need to be adopted by neurologists, in order to improve the interpretation of the results of published papers and to develop more proper outcome measures for use in future follow-up studies and clinical trials.

References

1. van Nes, S.I., C.G. Faber, and I.S. Merkies, *Outcome measures in immune-mediated neuropathies: the need to standardize their use and to understand the clinimetric essentials.* J Peripher Nerv Syst, 2008. **13**(2): p. 136-47.
2. Feinstein, A.R., *Clinimetrics*. 1987, New Haven and London: Yale University Press.
3. Hughes, R.A., et al., *Intramuscular interferon beta-1a in chronic inflammatory demyelinating polyradiculoneuropathy.* Neurology, 2010. **74**(8): p. 651-7.
4. van Schaik, I.N., et al., *Pulsed high-dose dexamethasone versus standard prednisolone treatment for chronic inflammatory demyelinating polyradiculoneuropathy (PREDICT study): a double-blind, randomised, controlled trial.* Lancet Neurol, 2010. **9**(3): p. 245-53.
5. Dalakas, M.C., et al., *Placebo-controlled trial of rituximab in IgM anti-myelin-associated glycoprotein antibody demyelinating neuropathy.* Ann Neurol, 2009. **65**(3): p. 286-93.
6. RMC trial group, *Randomised controlled trial of methotrexate for chronic inflammatory demyelinating polyradiculoneuropathy (RMC trial): a pilot, multicentre study.* Lancet Neurol, 2009. **8**(2): p. 158-64.
7. Hughes, R.A., et al., *Intravenous immune globulin (10% caprylate-chromatography purified) for the treatment of chronic inflammatory demyelinating polyradiculoneuropathy (ICE study): a randomised placebo-controlled trial.* Lancet Neurol, 2008. **7**(2): p. 136-44.
8. Leger, J.M., et al., *A randomized placebo-controlled trial of rituximab in IgM anti-myelin-associated glycoprotein antibody demyelinating neuropathy (RIMAG study).* J Peripher Nerv Syst, 2011. **16**(supplement 3): p. S73-S74.
9. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
10. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.
11. Rasch, G., *Probabilistic models for some intelligence and attainment tests.* 1960, Copenhagen: Danmarks Paedagogiske Institut.
12. Tennant, A. and P.G. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* Arthritis Rheum, 2007. **57**(8): p. 1358-62.
13. Altman, D.G. Practical statistics for medical research. 1991, London: Chapman and Hall.
14. Andrich, D., et al., *Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0).* 2003, Rumm Laboratory Pty Ltd.: Duncraig, Western Australia.
15. Fischer, W.P., *Reliability statistics.* Rasch Meas Trans, 1992. **6**: p. 238.
16. Linacre, J., *Sample size and item calibration stability.* Rasch Meas Trans, 1994. **7**: p. 328.
17. van Nes, S.I., et al., *Improving fatigue assessment in immune-mediated neuropathies: the modified Rasch-built fatigue severity scale.* J Peripher Nerv Syst, 2009. **14**(4): p. 268-78.
18. Ware Jr, J.E., et al., *SF-36 health survey Manual and interpretation guide*. 1993, The Health Institute, New England Medical Center: Boston.
19. Dorans, N.J. and P.W. Holland, *DIF detection and description: Mantel-Haenszel and standardisation*, in *Differential Item Functioning*, P.W. Holland and H. Wainer, Editors. 1993, Lawrence Erlbaum Associates: Hillsdale, NY. p. 36-66.

Chapter 3.2

Improving fatigue assessment in immune-mediated neuropathies: the Rasch-built modified fatigue severity scale

Sonja I. van Nes[1], Els K. Vanhoutte[2], Catharina G. Faber[2], Marcel Garssen[3], Pieter A. van Doorn[1], Ingemar S.J. Merkies [1,2,4], on behalf of the PeriNomS study group

[1]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of Neurology, Maastricht University Medical Centre, Maastricht, the Netherlands
[3]Department of Neurology, Jeroen Bosch Hospital, Den Bosch, the Netherlands
[4]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands

J Peripher Nerv Syst. 2009 Dec; 14(4): 268-78

Abstract

Fatigue is a major disabling complaint in patients with immune-mediated neuropathies. The 9-item fatigue severity scale (FSS) has been used to assess fatigue in these conditions, despite having limitations due to its classic ordinal construct. The aim was to improve fatigue assessment in patients with immune-mediated neuropathies through evaluation of the FSS using a modern clinimetric approach (Rasch unidimensional measurement models (RUMM2020)). Included were 192 stable patients with Guillain-Barré syndrome (GBS) , chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) or polyneuropathy associated with monoclonal gammopathy of undetermined significance (MGUSP).

The obtained FSS data were exposed to RUMM2020 model to investigate whether this scale would meet its expectations. Also, reliability and validity studies were performed. The original FSS did not meet the Rasch model expectations, primarily based on two misfitting items, one of these also showing bias towards the factor 'walking independent'. After removing these two items and collapsing the original 7-point Likert options to 4-point response categories for the remaining items, we succeeded in constructing a 7-item Rasch-built scale that fulfilled all requirements of unidimensionality, linearity, and rating scale model. Good reliability and validity were also obtained for the modified FSS. In conclusion, a 7-item linearly weighted Rasch-built modified FSS is presented for more proper assessment of fatigue in future studies in patients with immune-mediated neuropathies.

Introduction

Fatigue is considered a major disabling symptom in patients with immune-mediated neuropathies, such as Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), and polyneuropathy associated with monoclonal gammopathy of undetermined significance (MGUSP).[1] Despite an apparent good physical recovery, fatigue may lead to substantial decrement in quality of life expectations. The 9-item fatigue severity scale (FSS) has been used to capture the consequences of fatigue in patients with immune-mediated polyneuropathies and in many other chronic medical conditions. The FSS has demonstrated its simplicity, applicability, and various reliability and validity forms.[2, 3]
Parts of its construct validity were also based on factor analysis and its sumscore has been recommended for scale score analyses. However, the use of a scale like the FSS has been criticised because it is based on the classical test theory (CTT), which recruits items without investigating their relevance for patients at examination.[4] Patients are requested to complete all the items in CTT, even though some may be inappropriate for their level of ability. Moreover, the suggested sumscores in CTT assume that given differences in these scores have equal meaning. Some caution is warranted when changes in sumscores are considered, because the whole score may not equal the sum of the parts.[5] Also, analysing data using primarily factor analysis is prone to misleading results, since observations are non-linear which may generate illusory factors.[6]

More than 30 multi-item fatigue scales with good clinimetric properties have been developed and used for various chronic illnesses in the last decades.[7] However, a linearly weighted fatigue scale is preferred using modern scientific methods in patients with immune-mediated neuropathies. The Rasch technology is increasingly being recognised as a modern test method that could facilitate this.[8-10] This technology attempts to transform ordinal scores into linear measures and is based on a logical assumption: individuals with high ability to perform a task should have an increased probability, relative to individuals with low levels, of getting a better score on any item.[11]

The aim of this study is to improve fatigue assessment in patients with immune-mediated neuropathies. Therefore, we examined whether the original FSS scale fulfils model expectations using a Rasch measurement model. Deviation from model expectation will be examined to determine whether the FSS construct can be improved. Finally, reliability and validity aspects will be investigated.

Patients and methods

Patients

Data obtained from 192 patients (163 with GBS, 21 with CIDP, and 8 with MGUSP) were re-examined for the purposes of this study. Patients were recruited from the Erasmus Medical Centre Rotterdam databank and the Dutch GBS study group and have already contributed to the clinimetric studies of the Inflammatory Neuropathy Cause and Treatment (INCAT) group and a published randomised trial evaluating fatigue in GBS.[3, 12] The following inclusion criteria were applied: a stable clinical condition was mandatory and was defined as an unchanged physical or functional condition in the last 6 months before the start of the studies as declared by the patients to the best of their knowledge. Where possible, findings at physical examination at entry were compared with available earlier physical examination data to ascertain stability. Patients were excluded from participation if there was any concomitant disease or use of medication that might cause chronic fatigue. All selected patients declared to have experienced only mild and transient fatigue prior to their illness and still had residual symptoms or signs resulting from their illness (f-score ≥ 1).[13] The patients with GBS and CIDP met the international criteria for their illness.[14, 15] The diagnosis MGUSP was established after excluding other underlying causes for the gammopathy and polyneuropathy.[16]

Assessment scales

- The *fatigue severity scale* (FSS), a brief and simple self-assessed questionnaire, has demonstrated its scientific soundness in immune-mediated neuropathies.[1, 2] The FSS is a 9-item questionnaire with response categories ranging from 1 ('strongly disagree'), 2 ('mainly disagree'), 3 ('partially disagree'), 4 ('do not agree/disagree'), 5 ('partially agree'), 6 ('mainly agree'), to 7 ('strongly agree') for each inquiry. The sumscore of the nine inquiries ranges from 9 ('no signs of fatigue') to 63 ('most disabling fatigue').
- The *short-form fatigue scale (SFFS)*, with acceptable validity and reliability, was administered in order to evaluate the construct convergent validity of the analysed FSS. The score of this scale ranges from 4 ('no signs of fatigue') to a maximum of 28 points ('most severe fatigue').[1, 17]
- The *f-score (GBS disability score)* is a 7-point disability scale ranging from 0('no symptoms') to 6 ('death').[18]

Test procedure

FSS data were obtained with informed written consent and after approval of the medical ethical committee of the Erasmus Medical Centre. Participants were lucid and received instructions on how to complete the fatigue forms. These questionnaires were answered in random order. Patients were examined at our outpatient clinic. Scores were double-checked before departure of the patients. The FSS was mailed 2-4 weeks later to all patients (second sample) for a second assessment (test-retest).

General Statistics and Rasch measurement aspects

Descriptive statistics

Personal aspects like age, gender, type of inflammatory neuropathy, duration of illness, and GBS disability level (f-score) were collected. For Rasch analyses purposes, these personal factors were arbitrarily categorised as follows: age category (1: < 40 years, 2: 40-59 years, 3: ≥ 60 years), gender (0: female, 1: male), illness type (1: GBS, 2: CIDP, 3: MGUSP), polyneuropathy form (1: acute = GBS; 2: chronic = CIDP and MGUSP patients), duration category (1: < 2 years; 2: 2-5 years, 3: ≥ 5 years), and walking independent category (0: unable = f-score > 2; 1: able = f-score ≤ 2). The rationale for these categories was to investigate their possible influence on experiencing fatigue.[19]

Sample size consideration

For Rasch analyses purposes, a convenient sample size of 150 patients has been suggested to estimate item difficulty, with a confidence interval of 99% and item calibrations within +0.5 logits.[20] In the current study, a sample of 192 patients was available, thus expecting an acceptable degree of precision of the Rasch analysis.

Rasch measurement model

The Rasch model is based on the probabilistic Guttman procedure.[21] For the purposes of the current study, the obtained FSS data and the selected 'personal factors' were subjected to Rasch analysis using the RUMM2020 software.[11, 22, 23] The following applications of RUMM2020 were used to examine whether the data fit the model expectations and items or patients not fulfilling these requirements were removed or subjected to readjustments to fit the model:

- *Fit statistics:* Three overall fit statistics are considered. Two are item-person interaction statistics transformed to approximate a z-score. Therefore if the items and persons fit the model, a mean around zero and a standard deviation of 1 would be expected. A third fit statistic score is an item-trait interaction statistic reported as a chi-square, reflecting the property of invariance across the trait. A non-significant chi-square indicates that the hierarchical ordering of the items does not vary across the trait, thus fulfilling the required property of invariance.
- *Internal reliability studies:* Internal consistency reliability of the scale was determined with the person separation index (PSI) or Cronbach's alpha (the latter is only possible in case of no missing values) using the logit scores for each person. A value of ≥ 0.7 was considered consistent with the scale being able to differentiate between at least two groups of patients and is seen as the minimum requirement for measurement.[24]

- *Fit residuals:* Individual person- and item-fit statistics were examined, both as residuals (deviations of items and persons from expected model scores: residuals between ±2.5 are considered adequate fit) and using a chi-squared statistic (significance indicating misfit).
- *Threshold examination:* The term threshold refers to the point between two adjacent response categories where either response is equally probable. That is the point where, for example in case of the FSS, a threshold would be the point between two adjacent response categories for each item (e.g., between 1: strongly disagree and 2: mainly disagree; see figure 1A). One of the most common sources of item misfit concerns respondents' inconsistent use of these response options. This results in what is known as 'disordered thresholds': the failure of respondents to use the response categories in a manner consistent with the level of the trait being measured. If needed, response options were collapsed in order to improve overall fit to the model.
- *Local dependency:* Residual correlations between items within the same scale are a source of misfit. This local dependency occurs when a respondent's answer to one item automatically has a bearing upon the answer to another item. Local dependency was examined, since it affects the estimation of test information and item discrimination parameters, thereby inflating the final scale score in a particular direction.[25]
- *Item bias:* A scale should always work in the same way irrespective of which personal factor (e.g., gender) is being examined.[26] For example, if men and women with equal levels of disability for their inflammatory neuropathy respond systematically differently to a fatigue item, then this item would be considered to display differential item functioning (DIF), and would violate the requirement for unidimensionality. The obtained data were examined for DIF using statistical (analyses-of-variance) and graphical methods.
- *Model forms:* The unrestricted partial credit Rasch model was used in the current study. Finally, the likelihood ratio test was applied to examine whether the final obtained model would fulfil the rating scale model requirements.
- *Test for unidimensionality:* Unidimensionality is tested by allowing the factor loadings on the first residual to determine subsets of items and then testing (paired t-test) to see if the person estimate (logit of person 'ability' or, in this case 'degree of fatigue') derived from these subsets significantly differs from that obtained from all items. The absence of any meaningful pattern in the residuals will support the assumption of local independence and unidimensionality of the scale.[27]

External construct validity and reliability studies

The external construct validity of the final modified FSS was assessed through convergent validity with the SFFS (intraclass correlation coefficient reported).[17]

Also, as part of the final analyses, graphical test-retest validity and reliability studies were performed comparing items' hierarchy and patients' locations between the two samples of patients.[28] The obtained correlations were quantified by estimation of the intraclass correlation coefficient using a one-way random effects analysis-of-variance model for groups' comparison.

Software

RUMM2020 Rasch software was used and further statistical and graphical analyses were undertaken using Stata 10.0 for Windows XP (Stata Statistical Software: Release 10.0, TX: Stata Corporation 2007). A value of p = 0.0055 (0.05/9) was used throughout the analyses, based on Bonferroni multiple testing corrections.[29]

Results

General description of patients

A total of 192 patients with immune-mediated neuropathies have been included in this study. The basic characteristics of all patients are presented in table 1. A total of 183 patients returned the second FSS assessment (95.3%). Forty-five patients (23.4%) were not able to walk independently (f-score > 2).

Table 1. Basic characteristics of patients with inflammatory polyneuropathies

	first sample	second sample (for test-retest)
n	192	183
Age, years; mean (SD), range	52.9 (14.3), 17-84	52.3 (14.1), 17-84
Gender; n (%) females males	 94 (49) 98 (51)	 91 (49.7) 92 (50.3)
GBS; n (%) CIDP; n (%) MGUSP; n (%)	163 (84.9) 21 (10.9) 8 (4.2)	158 (86.3) 20 (10.9) 5 (2.7)
FSS mean (SD), range	52.1 (10.3), 9-63	50.2 (11.2), 9-63
F-score (GBS disability score); n (%) 1 2 3 4	 50 (26.0) 97 (50.5) 34 (17.7) 11 (5.7)	 46 (25.1) 95 (51.9) 31 (16.9) 11 (6.0)
Duration of symptoms, years, Mean (SD), range	5.5 (5.9), 0-41	5.2 (5.0), 0-28

Legend to table 1. GBS = Guillain-Barré syndrome; CIDP = chronic inflammatory demyelinating polyradiculoneuropathy; MGUSP = monoclonal gammopathy of undetermined significant related polyneuropathy; FSS = fatigue severity scale.

Initial Rasch studies on the FSS

Fit statistic description

The original 9-item FSS scale did not meet the Rasch model expectations. There was a substantial items misfit with a mean fit residual of -0.253 and a SD of 2.058. The patients demonstrated reasonable fit statistics (mean fit residual -0.350, SD 1.025). There was a significant item-trait chi-square probability (p=0.000006) indicating a highly significant deviation from the model. The Cronbach's alpha was 0.92 and the PSI 0.93. Individual item fit statistics were also investigated and presented in table 2. As can be seen, items 1 and 8 demonstrated a misfit due to fit residuals exceeding ±2.5 (item 1: +4.418 and item 8: -3.185). Item 1 had also a significant low chi-square probability. The high negative fit residual of item 8 demonstrates its redundancy to the data set. Item 8 is not adding any information to the data.

Table 2. Rasch fit statistics for the initial 9 items FSS data

FSS item	Location	SE	FitResid	DF	ChiSq	DF	Prob
1	0.134	0.07	4.418	148.11	26.828	2	0.000001
2	-0.635	0.096	0.363	148.11	5.002	2	0.081988
3	-0.067	0.073	0.382	148.11	0.784	2	0.675829
4	-0.043	0.078	-1.14	148.11	4.032	2	0.133195
5	0.892	0.068	-0.092	148.11	2.736	2	0.254641
6	-0.605	0.085	-0.894	148.11	4.229	2	0.120719
7	-0.128	0.082	-1.329	148.11	3.424	2	0.1805
8	0.085	0.069	-3.185	148.11	8.66	2	0.013165
9	0.367	0.068	-0.796	148.11	1.376	2	0.50254

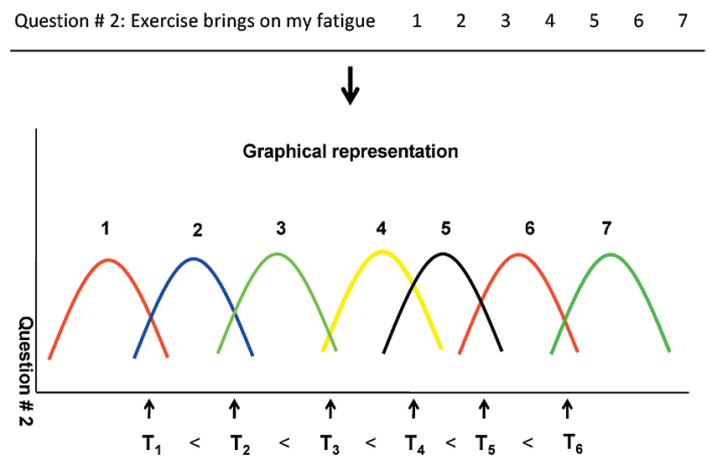
Legends to table 2. Items 1 and 8 demonstrated a misfit due to fit residuals exceeding ±2.5. Item 1 had also a significant low chi-square probability. The high positive labelled fit residual and significant probability of item 1 suggests that the response to this item differed from the responses to the remaining scale items. The high negative fit residual of item 8 demonstrates its redundancy to the data set; in other words, item 8 is not adding any information to the data.
FSS = fatigue severity scale; SE = standard error; FitResid = Fit residuals; DF = degrees of freedom; ChiSq = Chi square; prob = probability.

Threshold examination (category probability curves analyses)

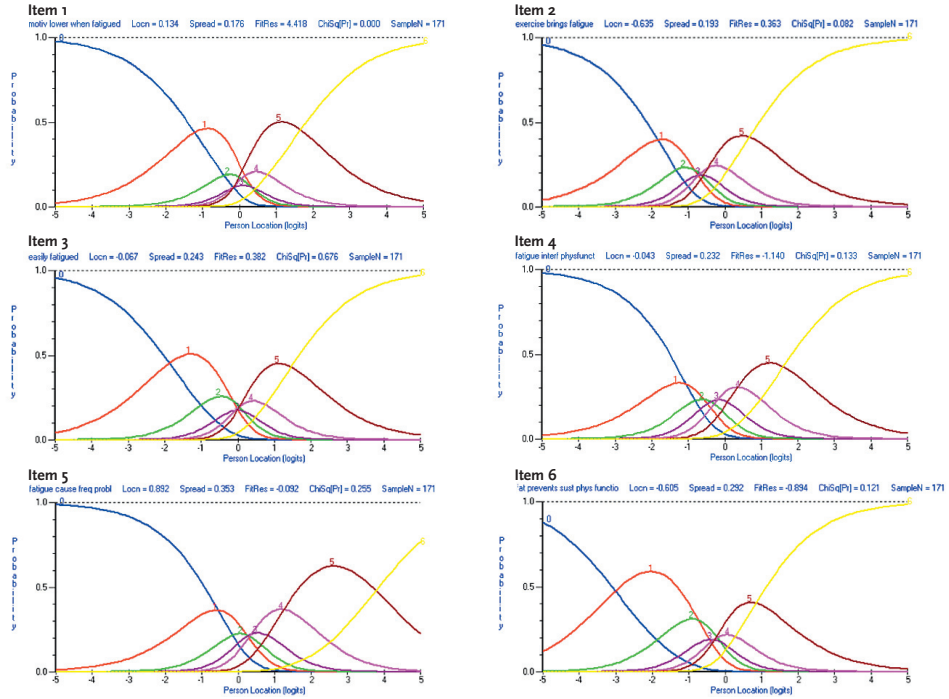
All items demonstrated in the mid response category area (response options ranging from 2: 'mainly disagree' up to 6: 'mainly agree') a general inability of the patients to differentiate between these options, with disordered thresholds for almost all items. Item 8 ('Fatigue is among my three most disabling symptoms') demonstrated the strongest disordered threshold pattern (figure 1B).

Figure 1. Category probability curves

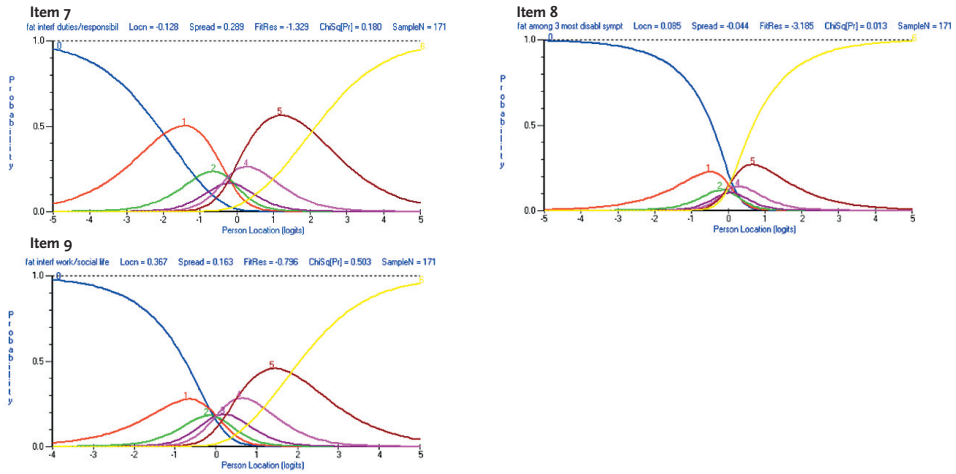
A. **Response options** 1=strongly disagree to 7=strongly agree



B.



B. Continued



Legend to figure 1.

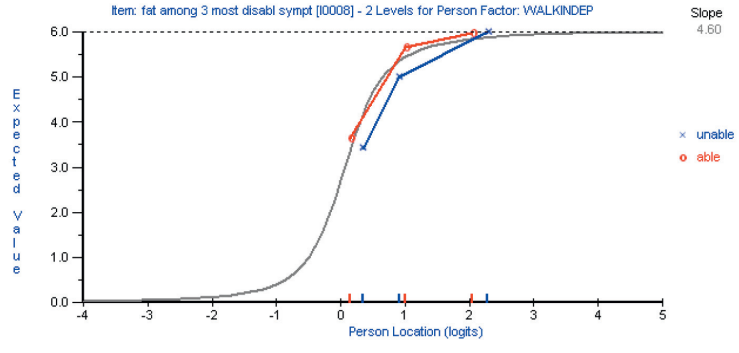
(A) Graphical respresentation of ideal ability of patients to discriminate between response options of fictitious items ranging from 1 ('strongly disagree') to 7 ('strongly agree'). FSS question number 2 is used as an example.

(B) Patients with immune-mediated neuropathies demonstrated unanimously inability to differentiate between the original response options of the 9-item FSS ranging from 2 ('mainly disagree') to 6 ('mainly agree'), leading to many inverted thresholds.

Local dependency (item residual correlations) and item bias analyses

No significant correlations were seen between the original nine items of the FSS. All items were subjected to DIF investigation for the selected personal factors. Only for item 8, a uniform DIF was demonstrated (p=0.000562; see figure 2). Item 8 was considered more difficult to answer for the 45 patients who could not walk independently. The remaining items did not show any item bias.

Figure 2. Item characteristic curve for item 8 related to personal factor 'ability to walk independent'



Legend to figure 2. Patients unable to walk independently demonstrated uniform DIF on item 8: this item was experienced as more difficult to answer compared with those able to walk independently.

Independent t-test

Based on the first principal components analysis, two item subsets were formed by grouping the three most positive loading items (item 1: loading 0.685; item 2: 0.661; item 6: 0.190) vs. the three most negative loading items (item 5: loading -0.557; item 9: -0.433; item 7: -0.390). The two subsets were compared through independent t-test and demonstrated a proportion of 0.0819, suggesting that the data did not show unidimensionality.

Data handling to improve the FSS fit to the Rasch model

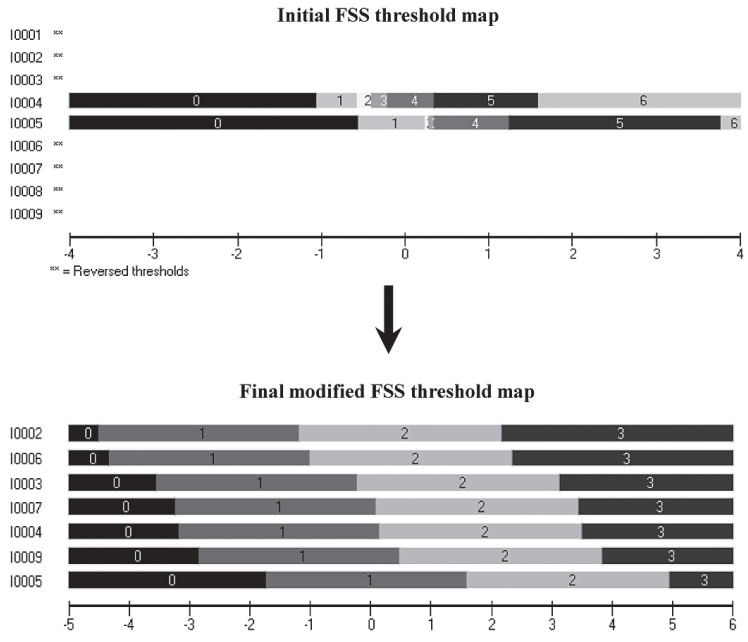
Rescoring items to improve thresholds and removing items

Based on the findings illustrated in figure 1, we decided to systematically (from item 1 to 9) rescore the items in order to restore reversed thresholds. Items 1, 2, 4, 5, 7, 8, and 9 were re-scored as follows: original scale scores 1 to 7, initial Rasch scores 0 - 6; rescoring: 0 = 0, 1 - 3 = 1; 4 – 5 = 2, and 6 = 3. Items 3 and 6 were rescored as follows after studying the location of disordered thresholds: 0 = 0, 1 - 2 = 1; 3 – 5 = 2, and 6 = 3. After performing these, items 1 and 8 still demonstrated misfit statistics and were subsequently removed from the analyses.

Threshold mapping, fit statistics and testing for rating scale model expectations

With the help of threshold mapping, we were able to restore the disordered thresholds and transformed the response categories in the remaining seven items to equivalent 4-point response categories (figure 3 and appendix). For the items and patients, mean fit residuals of -0.289 (SD: 1.113) and -0.275 (SD: 1.091) were obtained, respectively. The Cronbach's alpha and PSI remained good (0.89 for both) with an overall chi-square probability of 0.22. All individual items had good fit statistics (table 3). Item 2 ('exercise brings on my fatigue') turned out to be the easiest inquiry to answer while item 5 ('fatigue causes frequent problems for me') was considered the most difficult (table 3). A p-value of 0.10 was obtained for the modified 7-item FSS version by means of a likelihood ratio test, implicating that the items meet the Rasch rating scale model expectations.

Figure 3. Rescoring response categories



Legend to figure 3. After deleting misfitting items and rescoring the response options (from seven categories to four categories), we were able to construct a modified 7-item FSS scale fulfilling all Rasch model expectations.

Table 3. Rasch fit statistics for the final 7-item FSS data

FSS item	Location	SE	FitResid	DF	ChiSq	DF	Prob
2	-1.174	0.165	1.95	140.43	0.109	2	0.947176
6	-0.995	0.162	-1.271	140.43	3.082	2	0.2142
3	-0.208	0.153	-0.874	140.43	1.137	2	0.566385
7	0.11	0.15	-1.139	140.43	2.83	2	0.242876
4	0.158	0.15	-0.197	140.43	3.311	2	0.19102
9	0.502	0.148	0.721	140.43	4.186	2	0.12334
5	1.607	0.145	-1.16	140.43	2.991	2	0.224179

Legends to table 3. Items 1 and 8 were removed due to misfitting the model and the remaining 7 items met the model requirements. Loading on location indicates item difficulty in patients with immune-mediated neuropathies (most negative = easiest, most positive = most difficult item to answer) FSS = fatigue severity scale; SE = standard error; FitResid = fit residuals; DF = degrees of freedom; ChiSq= Chi square; prob = probability.

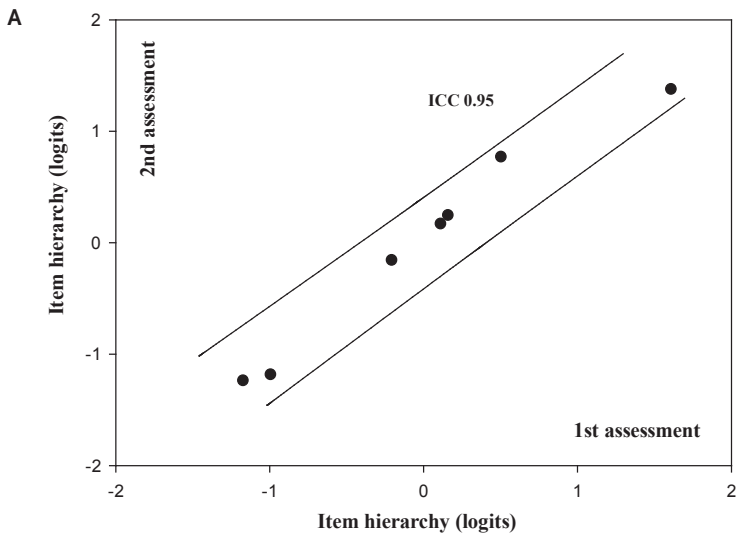
Item bias and test for unidimensionality

The individual items were all free from DIF. Based on the first principal components analysis, two item subsets were formed again: positive loading residual items (2, 3, and 6) vs. negative items (5, 7, and 9). The independent t-test between these two subsets of items demonstrated a proportion of 0.02, which shows unidimensionality. Finally, the raw 7-item FSS sumscore, ranging from 0 ('no signs of fatigue') to 21 ('most disabling fatigue') can be transformed to logits for the patients, ranging from -6 to +6. Only 1.1% and 9.7% of the patients sample size had a floor or ceiling effect, respectively.

External construct validity and reliability studies

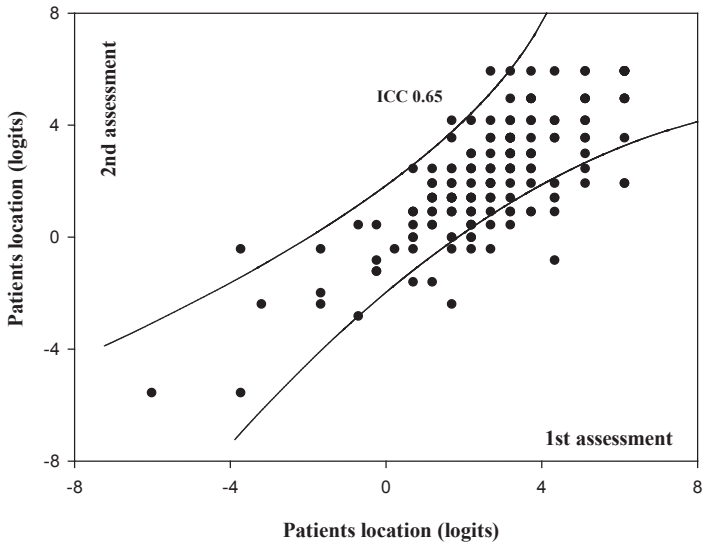
The modified FSS demonstrated an acceptable correlation with the SFFS (intraclass correlation coefficient: 0.71, p<0.0001). Test retest reliability studies were performed to investigate whether the items' hierarchy and patients' locations were consistent over time. As demonstrated in figure 4, all seven items were located within the 95% confidence interval lines, indicating ideal invariance. Most patients were also within the 95% confidence interval limits.

Figure 4. Test retest reliability studies between item hierarchy (A) and person location (B) for the Rasch-built modified FSS in immune-mediated neuropathies



Legend to figure 4.
(A) Test-retest reliability studies between items' hierarchy assessed at the first and second FSS assessment and the 95% confidence interval (solid lines) for the ideal invariance. As can be seen, all items were ordered in a hierarchal way, thus demonstrating ideal invariance (intraclass correlation coefficient for the item estimates of 0.95).

B



Legend to figure 4.
(B) Test-retest reliability studies between persons' location assessed at the first and second FSS assessment and the 95% confidence interval (solid lines) for the ideal invariance. Acceptable reproducibility was obtained between the patients' locations (intraclass correlation coefficient for the item estimates of 0.75).

Discussion

To our knowledge, the current study is the first study examining the modern clinimetric aspects of the FSS in patients with immune-mediate neuropathies. The original 9-item FSS failed in meeting the requirements addressed by the Rasch model.[9] However, after systematic evaluation and readjustment of the data, we were able to present a linearly weighted 7-item (4 response categories for each item) Rasch-built modified FSS, that fulfilled all clinimetric requirements, including validity and reliability (see also appendix).[10, 11, 28] The patients' locations were acceptable, although most patients were within the 95% confidence interval limits and had high scores which indicates having a more impaired status due to severe fatigue (see figure 4B). This is in conformity with earlier findings.[1] An elegant and equivalent study on FSS has been recently presented, that demonstrated quite similar findings in patients with multiple sclerosis (MS).[30] In this paper, more items were disordered and more DIF was seen in patients at examination. The final modified FSS scale in patients with MS turned out to be somewhat different. It is conceivable that the obtained differences in experiencing fatigue between patients with MS versus those in our study could be related to differences in possible confounding factors such as location of illnesses (central versus peripheral nervous system), differences in disability, and possible differences in pathophysiological mechanism of fatigue in these disorders.

In the current study, the most striking pattern was the patients' unanimous inability to differentiate between the original response categories ranging from 2 to 6, leading to many inverted thresholds (figures 1 and 3). Although the authors of the original FSS attempted to improve differentiation between patients, the amount of response categories led to confusion in patients with immune-mediated neuropathies that could only be visualised using a modern technique like the Rasch. This was also in conformity with earlier reports demonstrating difficulties of patients to discriminate among more than three response categories.[31]

The obtained 7-item Rasch-built modified FSS has the advantages of a modern test theory (e.g., linearity, unidimensionality) and its use is believed to improve fatigue assessment in immune-mediated neuropathies and will certainly help clinicians to focus on ameliorating the enormous disabling impact of fatigue. Therefore, we suggest the use of this modified 7-item FSS in future studies targeting fatigue in patients with GBS, CIDP, and MGUSP.

Appendix

Transforming original 9-item FSS to Rasch-built modified 7-item FSS

Original fatigue severity scale								
1=strongly disagree; 2=mainly disagree; 3=partially disagree; 4=do not agree/disagree; 5=partially agree; 6=mainly agree 7=strongly agree (circle one answer per question)								
1	My motivation is lower when I am fatigued	1	2	3	4	5	6	7
2	Exercise brings on my fatigue	1	2	3	4	5	6	7
3	I am easily fatigued	1	2	3	4	5	6	7
4	Fatigue interferes with my physical functioning	1	2	3	4	5	6	7
5	Fatigue causes frequent problems for me	1	2	3	4	5	6	7
6	My fatigue prevents sustained physical functioning	1	2	3	4	5	6	7
7	Fatigue interferes with carrying out certain duties and responsibilities	1	2	3	4	5	6	7
8	Fatigue is among my three most disabling symptoms	1	2	3	4	5	6	7
9	Fatigue interferes with my work, family, or social life	1	2	3	4	5	6	7

↓

Rasch-built modified fatigue severity scale					
The higher the score you choose, the more you agree with the question (the lower the score the less you agree)		disagree			agree
1	Exercise brings on my fatigue	0	1	2	3
2	I am easily fatigued	0	1	2	3
3	Fatigue interferes with my physical functioning	0	1	2	3
4	Fatigue causes frequent problems for me	0	1	2	3
5	My fatigue prevents sustained physical functioning	0	1	2	3
6	Fatigue interferes with carrying out certain duties and responsibilities	0	1	2	3
7	Fatigue interferes with my work, family, or social life	0	1	2	3

Score range of the Rasch-built modified fatigue severity scale: 0 ('no signs of fatigue') to 21 ('most disabling fatigue').

References

1. Merkies, I.S., et al., *Fatigue in immune-mediated polyneuropathies. European Inflammatory Neuropathy Cause and Treatment (INCAT) Group.* Neurology, 1999. **53**(8): p. 1648-54.
2. Krupp, L.B., et al., *The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus.* Arch Neurol, 1989. **46**(10): p. 1121-3.
3. Merkies, I.S.J., *Evaluation of scales and measurement instruments in immune-mediated polyneuropathies.* 2001, Erasmus Medical Centre: Rotterdam.
4. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
5. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.
6. Linacre, J.M., *Rasch First or Factor First? 11:4 p. 603.* Rasch Measure Trans, 1998. **11**(4): p. 603.
7. Dittner, A.J., S.C. Wessely, and R.G. Brown, *The assessment of fatigue: a practical guide for clinicians and researchers.* J Psychosom Res, 2004. **56**(2): p. 157-70.
8. Bond, T.G. and C.M. Fox, *Applying the Rasch model: fundamental measurement for the human sciences.* 2001, New York: Lawrence Erlbaum Associates.
9. Rasch, G., *Probabilistic models for some intelligence and attainments tests*, ed. U.o.C. Press. 1980, Chicago.
10. Streiner, D.L. and G.R. Norman, *Health measurement scales. A practical guide to their development and use*. 1998, New York: Oxford University Press, 2nd ed.
11. Pallant, J.F. and A. Tennant, *An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS)*. Br J Clin Psychol, 2007. **46**(Pt 1): p. 1-18.
12. Garssen, M.P., et al., *Amantadine for treatment of fatigue in Guillain-Barre syndrome: a randomised, double blind, placebo controlled, crossover trial.* J Neurol Neurosurg Psychiatry, 2006. **77**(1): p. 61-5.
13. Kleyweg, R.P., F.G. van der Meche, and P.I. Schmitz, *Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barre syndrome.* Muscle Nerve, 1991. **14**(11): p. 1103-9.
14. *Research criteria for diagnosis of chronic inflammatory demyelinating polyneuropathy (CIDP). Report from an Ad Hoc Subcommittee of the American Academy of Neurology AIDS Task Force.* Neurology, 1991. **41**(5): p. 617-8.
15. Asbury, A.K. and D.R. Cornblath, *Assessment of current diagnostic criteria for Guillain-Barre syndrome.* Ann Neurol, 1990. **27 Suppl**: p. S21-4.
16. Miescher, G.C. and A.J. Steck, *Paraproteinaemic neuropathies.* Baillieres Clin Neurol, 1996. **5**(1): p. 219-32.
17. Alberts, M., et al., *'Abbreviated fatigue questionnaire': a practical tool in the classification of fatigue.* Ned Tijdschr Geneeskd, 1997. **141**(31): p. 1526-30.
18. Hughes, R.A., et al., *Controlled trial prednisolone in acute polyneuropathy.* Lancet, 1978. **2**(8093): p. 750-3.
19. Jonker, A.A., et al., *Persistent Deterioration of Functioning (PDF) and change in well-being in older persons.* Aging Clin Exp Res, 2008. **20**(5): p. 461-8.
20. Linacre, J.M., *Sample size and item calibration stability.* Rasch Measure Trans, 1994. **7**: p. 28.
21. Altman, D.G., *Practical statistics for medical research*, ed. C. Hall. 1990, London.
22. Andrich, D., et al., *Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0)*, 2003, Rumm Laboratory Pty Ltd.: Duncraig, Western Australia.
23. Smith, R.M., *Fit analysis in latent trait measurement models*, J Appl Meas, 2000. **1**: p. 199-218..
24. Fisher, W.P., *Reliability Statistics.* Rasch Measure Trans, 1992. **6**: p. 238.
25. Embretson, S.E. and S.P. Reise, *Item response theory for psychologists.*, ed. L.E. Associates. 2000, New Jersey.
26. Dorans, N.J. and P.W. Holland, *DIF detection and description: Mantel_Haenszel and standadisation*, in *Differential item functioning*, P.W. Holland and H. Wainer, Editors. 1993, Lawrence Erlbaum Associates: Hillsdale, NY. p. 36-66.
27. Smith, E.V., Jr., *Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals.* J Appl Meas, 2002. **3**(2): p. 205-31.
28. Wright, B.D. and M.H. Stone, *Best test design*. 1979, Chicago: Mesa Press.

29. Bland, J.M. and D.G. Altman, *Multiple significance tests: the Bonferroni method.* Bmj, 1995. **310**(6973): p. 170.
30. Mills, R., et al., *Rasch analysis of the Fatigue Severity Scale in multiple sclerosis.* Mult Scler, 2009. **15**(1): p. 81-7.
31. Vandervelde, L., *Activity limitations in patients with neuromuscular disorders*. 2008, Universite catholique de Louvain.

Chapter 3.3

Rasch-built overall disability scale (R-ODS) for immune-mediated peripheral neuropathies

Sonja I. van Nes[1], Els K. Vanhoutte[2], Pieter A. van Doorn[1], Mieke Hermans[2], Mayienne Bakkers[2], Krista Kuitwaard[1], Catharina G. Faber[2], Ingemar S.J. Merkies[1,2,3], on behalf of the PeriNomS study group

[1]Department of neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of neurology, Maastricht University Medical Centre, Maastricht, the Netherlands
[3]Department of neurology, Spaarne Hospital, Hoofddorp, the Netherlands

Neurology 2011;76:337-345

Abstract

Objective: To develop a patient-based, linearly weighted scale that captures activity and social participation limitations in patients with Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) and monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP).

Methods: A preliminary Rasch-built overall disability scale (R-ODS) containing 146 activity and participation items was constructed, based on the WHO international classification of Functioning, Disability and Health, literature search, and patient interviews. The preliminary R-ODS was assessed twice (interval: 2-4 weeks; test-retest reliability studies) in 294 patients who experienced GBS in the past (n=174) or currently have stable CIDP (n=80) or MGUSP (n=40). Data were analysed using the Rasch unidimensional measurement models (RUMM2020).

Results: The preliminary R-ODS did not meet the Rasch model expectations. Based on disordered thresholds, misfit statistics, item bias, and local dependency, items were systematically removed to improve the model fit, regularly controlling the class intervals and model statistics. Finally, we succeeded in constructing a 24-item scale that fulfilled all Rasch requirements. 'Reading a newspaper/book' and 'eating' were the two easiest items, 'standing for hours' and 'running' were the most difficult ones. Good validity and reliability were obtained.

Conclusion: The R-ODS is a linearly weighted scale that specifically captures activity and social participation limitations in patients with GBS, CIDP and MGUSP. Compared to the overall disability sumscore (ODSS), the R-ODS represents a wider range of item difficulties thereby better targeting patients with different ability levels. If responsive, the R-ODS will be valuable for future clinical trials and follow-up studies in these conditions.

Introduction

Disability has been proposed as the preferential level for measuring therapeutic response in immune-mediated neuropathies.[1] However, most disability scales used in these disorders are based on the classical test theory (CTT).[2-7] A summary of the disadvantages of CTT has been provided.[2] Generally, patients are requested to complete all items of CTT-based scales, even though some may be irrelevant for their level of ability. Furthermore, a sum of item scores is often calculated assuming equal relevance and hence weighting of each item which is highly unlikely.[2, 8] Given these limitations, a modern approach like the Rasch technology is needed to develop a scale measuring disability (i.e., activity and social participation limitations) in patients with immune-mediated neuropathies.[9, 10] Rasch models the probability that a person will be able to complete an item, only dependent on the item difficulty and the person's level of ability.[9, 11, 12] It enables scales measuring the same health construct to be equated on the same linear ruler optimising comparisons of studies, meta-analyses, and systematic reviews.[10]

The AMC linear disability score (ALDS) item bank and ACTIVLIM are modern disability scales.[13-15] However, their calibration may not be representative to patients with immune-mediated neuropathies since these scales are not disease-specific. Therefore, we developed a linearly weighted scale at the activity and participation level, the Rasch-built overall disability scale (R-ODS), specifically for patients with Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyneuropathy (CIDP) and monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP) and evaluated its validity and reliability aspects.

Methods

Participants

A total of 852 members of the Dutch society of neuromuscular disorders (Vereniging Spierziekten Nederland [VSN]), registered as having a polyneuropathy were initially requested to participate. For this study, we used data from patients who have been diagnosed with GBS, CIDP or MGUSP (254 of the 511 members who returned their questionnaires). Additionally, 40 patients (11 GBS, 22 CIDP and 7 MGUSP) were recruited at the university outpatient clinics of Rotterdam and Maastricht, in the Netherlands. All participants were recruited between January 2007 and July 2009. Participants aged 18 years and older, with a stable clinical condition, defined as an unchanged physical and functional condition over 2 months prior to the study as declared by the participant, were included.

Standard protocol approvals, registrations, and patient consents

The local medical ethics committee approved the protocol. Written informed consent was obtained from all participants.

Questionnaire development

Step I: International development procedures were applied to construct the preliminary R-ODS.[12, 16, 17] Since we aimed to develop an outcome measure at the activity and participation level, we critically reviewed all potential items listed under these headings using the International Classification of Functioning, Disability and Health (ICF).[18] The chapters learning and applying knowledge, general tasks and demands, and communication were excluded. We selected defined items from this ICF list with probable relevance to polyneuropathy patients based on known clinical characteristics. Efforts were made to describe these items in a concise, simple and unambiguous way.

Step II: A systematic Medline and Embase search was performed for literature published over the last four decades, reviewing all scales at the activity and participation levels in patients with any form of (poly)neuropathy. Reports published in English were identified with reference tracing using the keywords: GBS, CIDP, acquired/idiopathic (poly)(radiculo)neuropathy, polyneuritis, gammopathy, dysimmune, paraprotein(a)emia, MGUSP, disability (scale), activity limitation, and handicap. Various handbooks of neurological rating scales were also evaluated. Additionally, items were selected from an earlier semi-structured interview focusing on daily activities addressed by patients with inflammatory polyneuropathies.[5] Eventually, we succeeded in constructing the preliminary R-ODS containing 146 items with 3 response options: 0 ('impossible to perform'), 1 ('performed with difficulty'), and 2 ('easily performed'). If an item was not applicable, the patient was requested to answer this item with the option 3 ('not applicable').

Step III: The selected items were judged by four neuromuscular experts. Based on the clinical characteristics of GBS, CIDP and MGUSP, items with insufficient face and content validity were removed.[16]

Procedures

Patients received the preliminary R-ODS plus the overall disability sumscore (ODSS) with instructions by mail.[6] They were requested to answer all questions by themselves and to answer 'impossible to perform' when unable to complete an item or 'able to perform, but with difficulty' when special devices or other forms of assistance were needed. After a period of 2-4 weeks, all participants were again requested to answer all questions of the preliminary R-ODS (test-retest reliability studies).

Rasch analyses

In the model construction, items scored as 3 ('not applicable') were interpreted as missing data. Items with > 10% missing values and patients with > 10% unanswered items were omitted as a quality control procedure. Thereafter, the remaining responses of the preliminary R-ODS were analysed using Rasch unidimensional measurement models (RUMM2020), which is based on a mathematical model proposed by Rasch, a Danish mathematician.[19] The finally constructed scale should be unidimensional, free from item bias, and without disordered thresholds or local dependency.[10, 20] We checked for possible item bias using the following personal factors, categorised arbitrarily as follows: age (<60 years vs. ≥60 years), gender (men vs. women), diagnosis (GBS vs. CIDP vs. MGUSP), diagnosis category (acute [GBS] vs. chronic [CIDP and MGUSP]), and duration of complaints (<2 years vs. 2-5 years vs. ≥5 years). Throughout the analyses, we continuously monitored the distribution of persons within the class intervals, the fit statistics and the independent tests for unidimensionality of the scale. Items and persons not fulfilling Rasch model criteria were evaluated and removed one by one if needed.

A sample size of approximately 250 is needed to obtain a 99% confidence with a stable item calibration within ± 0.5 logits, hence providing a stable model.[21] We expected an acceptable degree of precision of the Rasch analyses, since a sample of 294 patients was available.

Validity

Four neuromuscular experts stated that the items of the final R-ODS were ordered as expected. The external construct validity was assessed by correlation with the ODSS (intraclass correlation coefficient after applying quantile regression analysis). To compare the difficulty range of the items of the ODSS and the final R-ODS in relation to the ability range of the patients in this population we used the anchor-based approach in RUMM2020 (targeting validity study). Two identical items in the ODSS and the R-ODS ('turning a key in a lock' and 'dressing upper body') were used as anchors to place both measures and all patients on the same linear ruler.

Reliability

Internal reliability was examined by determining the person separation index (PSI). In general, a PSI above 0.7 is considered as acceptable, indicating the ability to identify at least two groups of patients.[22] Moreover, test-retest reliability studies were performed to investigate whether hierarchy of item difficulty and patient ability location were consistent over time.[23] Reliability was quantified by calculation of the intraclass correlation coefficient using a one-way random effects analysis of variance (ANOVA) model for group comparison.

Statistics and software

Rasch analyses were performed with the partial credit model as default, using RUMM2020 software. Further analyses were undertaken using Stata 10.0 for Windows XP. The p-value was adjusted throughout the analyses, based on Bonferroni multiple testing corrections.[24]

Results

General description of patients

A total of 511 of the 852 approached members of the Dutch society of neuromuscular disorders returned the first questionnaires (response rate of 60%). For the current study, we only used data of patients with a clinically stable form of immune-mediated polyneuropathy (174 GBS, 80 CIDP and 40 MGUSP). Most patients (74%) reported walking difficulties, 47% mentioned at least moderate symptoms in their upper limbs. A total of 255 patients completed the second preliminary R-ODS assessment (87%). The basic characteristics of all participants are presented in table 1.

Table 1. Basic characteristics of study population (first sample, n=294)

Diagnosis; n (%)	
GBS	174 (59)
CIDP	80 (27)
MGUSP	40 (14)
Age in years; mean (SD), range	60.2 (13,5), 18-91
Gender; n (%)	
Women	132 (45)
Men	162 (55)
Duration of symptoms, years; mean (SD), range	8.2 (7.8), 0.5 - 40
ODSS arm grade; n (%)*	
≤ 1 (normal or minor symptoms, not affecting any)	123 (53.3)
2 (moderate symptoms, not preventing any)	87 (37.7)
3 (severe symptoms, preventing at least one not all)	20 (8.7)
≥ 4 (severe symptoms, preventing all or no purposeful movements)	1 (0.4)
ODSS leg grade; n (%)*	
0 (normal)	59 (25.5)
1-2 (walking affected, with or without gait disturbance)	103 (44.6)
3 (unilateral support)	40 (17.3)
4 (bilateral support)	10 (4.3)
≥ 5 (wheelchair)	19 (8.2)

* Corresponding ODSS grades were only collected in 231 of the 294 patients

Legend to table 1. GBS = Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) and monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP).

Data quality control (Step 0)

Based on experts' opinion, n=12 items with insufficient face and content validity were removed (questionnaire development - step III). Also, another n=36 items with >10% missing values and n=14 patients with >10% unanswered items were omitted as a quality control procedure.

Initial Rasch analysis on the preliminary R-ODS

The first Rasch analysis was performed on the remaining 98 items. The mean items' fit residual showed reasonable model fit (mean -0.188, SD 1.119). However, the mean persons fit residual was -0.363 with a standard deviation exceeding the expected value of 1 (SD 1.428). The significant Chi-square probability implied no invariance of items. Initially, a proportion of 0.20 (95% confidence interval: 0.18-0.23) of the t-tests performed fell outside the ±1.96 range, indicating multidimensionality.

Stepwise approach to fit data to the Rasch model

Step 1. Disordered thresholds were seen in two items ('driving a car' and 'riding a bike'). Apparently, the response categories were conflicting for the patients. These two items were removed.

Step 2. Individual item and person fit statistics were inspected. Individual item fit statistics of 19 items demonstrated misfit to the model (fit residuals exceeding ± 2.5 or a significant chi-square probability or both). These 19 items were removed one by one, starting with the item with the highest deviation. In addition, two patients showed misfit (fit residuals exceeding ± 2.5) and were also removed.

Step 3. In six items (e.g., 'standing up from squat position', 'shaving/epilating', and 'vacuum cleaning') men and women with equal ability levels responded in a significant different manner (demonstrating uniform differential item functioning (DIF) on gender). Another three items demonstrated item bias on personal factor diagnosis (category) (two items had uniform DIF, one item non-uniform DIF). Finally, one item demonstrated item bias (uniform DIF) on the duration of complaints. All these items were removed one by one.

Step 4. A systematic evaluation of the correlation matrix findings (starting with the highest correlations ≥0.7, then ≥0.6, through to ≥0.28) was performed to identify local dependency (e.g., a person answering positively to the item 'able to walk 1 km' will most likely also respond positively on the item 'able to walk 100 m'). Item characteristic curves of each correlating item set were subsequently inspected to select the items that best fit the expected model curve. The item showing the most over- or under-discrimination was then removed. Eventually, a total of 43 items were stepwise removed.

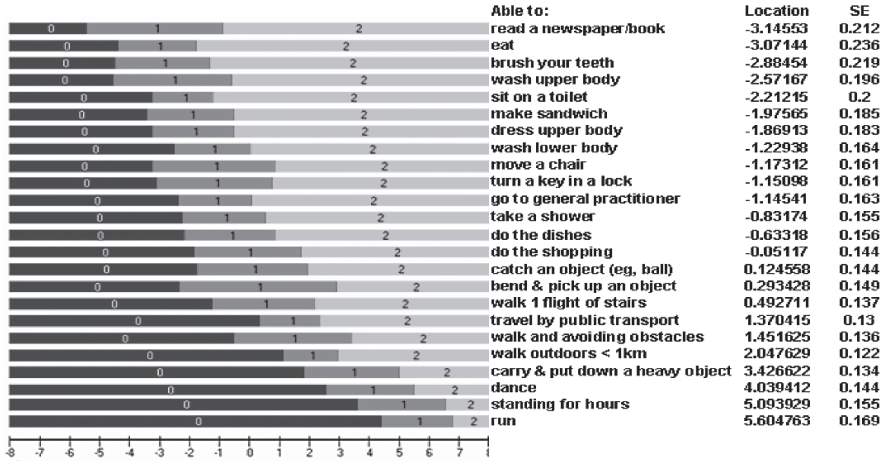
Final R-ODS

After completing these procedures, we succeeded in constructing a 24-tem scale which met the Rasch model expectations, the final R-ODS. Mean fit residuals for

items were -0.238 (SD 0.853) and for persons 0.352 (SD 0.839). The chi-square probability was p=0.29, thereby supporting the presence of invariance of item difficulty across the scale. Based on the first principal components analysis, two subsets of items were formed, with the six most positively loaded vs. the six most negatively loaded items. The independent t-tests between these two subsets were significant at the 5% level in 14 of the 261 patients (0.053%, confidence interval 0.027-0.080%) which indicates acceptable unidimensionality.

In the final R-ODS scale, the item 'reading a newspaper/book' was the easiest to perform while 'running' turned out to be the most difficult task (figure 1). Item difficulty ranges from -3.15 to 5.60 logits and patient ability level from -6.95 to 8.11 logits. Table 2 provides a nomogram allowing the translation of raw sumscores of the final R-ODS (range 0 to 48) to logits, placing patients' estimates on the same log-odds units (logit) scale. Since logits are difficult to interpret instinctively, we have converted the person locations into a centile metric score with values ranging from 0 (most severe activity and social participation limitations) to 100 (no activity and social participation limitations).

Figure 1. Threshold map and overall item difficulty locations for each item of the final Rasch-built overall disability scale (R-ODS)



Legend to figure 1. Black sections (0) = 'impossible to perform', dark grey sections (1) = 'with difficulty performed', and light grey sections (2) = 'easily performed'. The length of the coloured sections at the left panel respresent the number of patients giving a certain response. A gradual shift is seen: the black section of the item 'reading a newspaper/book' is much shorter than the black section of the item 'running'. Patients considered 'reading a newspaper/book' the easiest item and 'running' the most difficult item. The right panel presents the overall item difficulty locations and corresponding standard errors for each item.

Table 2. Nomogram

R-ODS summed raw score	Rasch person location (logits)	centile metric
0	-6.95	0
1	-6.03	6
2	-5.36	11
3	-4.87	14
4	-4.48	16
5	-4.14	19
6	-3.84	21
7	-3.57	22
8	-3.32	24
9	-3.09	26
10	-2.87	27
11	-2.66	28
12	-2.46	30
13	-2.26	31
14	-2.07	32
15	-1.88	34
16	-1.70	35
17	-1.52	36
18	-1.33	37
19	-1.15	39
20	-0.97	40
21	-0.79	41
22	-0.61	42
23	-0.42	43
24	-0.24	45
25	-0.05	46
26	0.14	47
27	0.34	48
28	0.53	50
29	0.73	51
30	0.94	52
31	1.15	54
32	1.36	55
33	1.58	57
34	1.81	58
35	2.04	60
36	2.28	61
37	2.54	63
38	2.80	65
39	3.09	67
40	3.40	69
41	3.74	71
42	4.11	73
43	4.54	76
44	5.03	80
45	5.59	83
46	6.25	88
47	7.07	93
48	8.11	100

Legend to table 2. This nomogram allows the translation of raw sumscores of the final R-ODS (range 0 to 48) to logits or to a centile metric score with values ranging from 0 (most severe activity and social participation restrictions) to 100 (no activity and social participation limitations). The corresponding logits in relation to the raw summed scores are provided by the RUMM software.

Validity

The intraclass correlation coefficient between the final R-ODS and the ODSS was 0.85, indicating good external construct validity. Furthermore, the population of patients examined demonstrated a ceiling effect of 5.8% on the final R-ODS vs. 19.4% ceiling effect on the ordinal ODSS (Student t-test p<0.0001). As shown in figure 2 item difficulties of the ODSS ranged from -1.869 to 1.336 logits (span of 3.205 logits). By contrast, the final R-ODS demonstrated a wider range of item difficulties (span of 8.750 logits).

Figure 2. Person locations vs. item locations

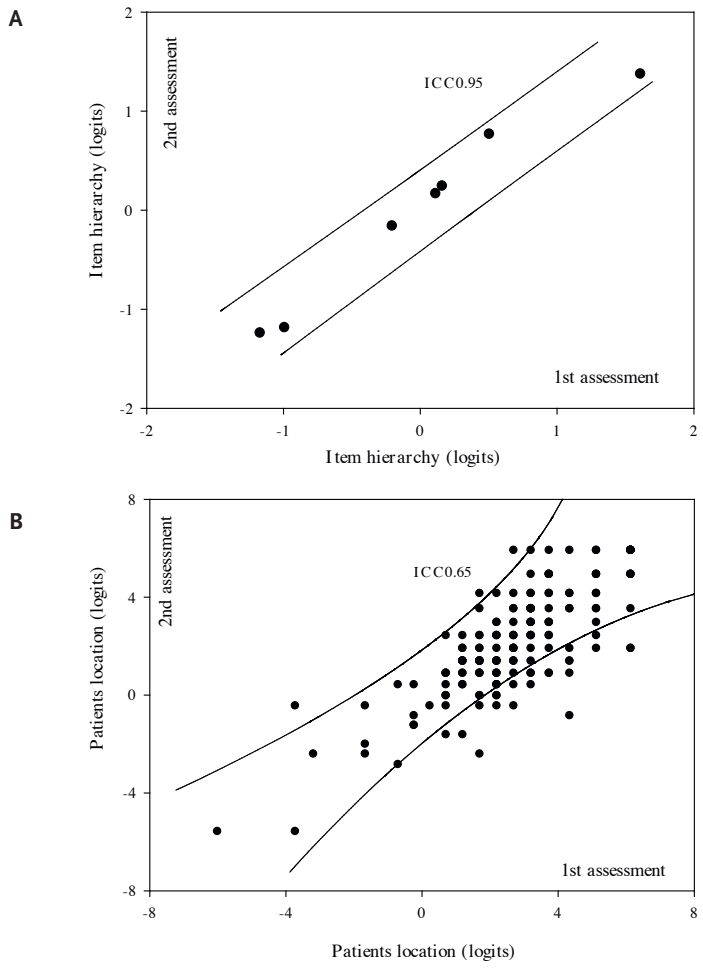
Persons	Location (logits)	ODSS items	R-ODS items
oooooooooooooo	8		
o			
ooooooooo	7		
oooooooo	6		
ooo			
			run
oooooooo			
oooo	5		stand for hours
o			
ooooooooo			
	4		dance
ooooooooooooooo			
o			
oooooooo			
oo			carry and put down a heavy object
ooooooooooooo			
ooooooooooo	3		
ooooooooooooo			
o			
ooooooooo			
oooooooo			
ooooooooo	2		walk outdoor < 1 km
ooooooooo			
ooo			
			walk and avoiding obstacles
ooooooo			
oooooooo		10 meters walk	travel by public transport
ooooooo	1		
oooooooooooooo			
oooooo			
			walk 1 flight of stairs
oooooo			
		handle buttons/ zippers	bend and pick up an object
ooooooo			
oooo	0	walking problems	catch an object (e.g., ball)
oooooo			do the shopping
ooooo			
ooooooo			
		wheelchair use	do the dishes
oooooo			
oo	-1	wash and brush hair	take a shower
ooooo		turn a key in a lock	go to general practitioner/turn a key in a lock/move a chair
o			wash lower body
oooo		use knife/ fork (spoon)	
ooo	-2	dress upper body	dress upper body / make a sandwich
ooo			
			sit on a toilet
o			
ooo			wash upper body
	-3		brush your teeth
o			eat / read a newspaper/ book
o			
o	-4		
oo			
o	-5		
o			
	-6		

Legend to figure 2. The numbers 8 to -6 under 'location (logits)' represent corresponding logits on a linear ruler. The open dots under 'persons' respresent the number of patients having a person location (ability level) at this point on the ruler. Each open dot corresponds to the location of 1 patient on the ruler. At the right side of the ruler, item locations (item difficulty) of the overall disability sumscore (ODSS) and the final Rasch-built overall disability scale (R-ODS) are presented enabling comparison through targeting (e.g., item difficulty range in relation to the different levels of ability in the sample).

Reliability

Internal reliability remained good indicated by a PSI of 0.97 for the final R-ODS. Also, test-retest reliability was good: item hierarchy and patients' locations were located within the 95% CI lines, reflecting good invariance (figure 3A and 3B).

Figure 3. Test-retest reliability of the final Rasch-built overall disability scale



Legend to figure 3. (A) Item difficulty hierarchy of the final R-ODS in the first vs. the second assessment. All items (dots) were located within the 95% confidence interval (solid lines) reflecting ideal reliability. (B) Patient's location in the first assessment vs. the second assessment. Almost all persons (dots) were located within the 95% confidence interval (solid lines). ICC= intraclass correlation coefficient

Discussion

We developed the R-ODS, a linearly weighted outcome measure that specifically captures activity and social participation limitations in patients with GBS, CIDP and MGUSP. This questionnaire was constructed using patient's perception of their ability to perform daily and social activities. All 24 items of the final R-ODS met the Rasch model expectations. Although the clinical characteristics of clinically stable GBS, CIDP and MGUSP patients are slightly different, their response pattern was quite similar in this study. Only three items of the preliminary scale were removed due to item bias for the personal factor diagnosis.

Since calibration of the R-ODS is based on the characteristics of our study population, we also questioned whether this population represents patients included in clinical trials. The baseline characteristics of patients in trials often report a sumscore of the ordinal ODSS of around 4. Moreover, inclusion criteria sometimes include minimum sumscores of the ODSS (e.g., at least two in the randomised methotrexate CIDP (RMC) trial, evaluating the efficacy of methotrexate in CIDP patients requiring immunoglobulin or corticosteroids) or a minimum ODSS arm or leg grade (e.g., at least arm grade 2 or leg grade 1 in a trial evaluating the efficacy of intravenous immunoglobulin in MGUSP).[25-27] The basic characteristics of our sample show arm grades ≥2 in 47% and leg grades ≥1 in 74% of the patients. This implies sufficient disability in the sample to calibrate the R-ODS for future use in CIDP and MGUSP trials. Only 30% of our sample was unable to walk independently (table 1, ODSS leg grade ≥ 3), corresponding to a GBS disability score of ≥ 3 which is often used as criterion to enrol patients in GBS trials.[28, 29] Clearly, patients with GBS in the acute phase may often have more severe symptoms, signs, and functional deficit than our clinically stable patients who experienced GBS in the past. It is therefore conceivable that the weights of the items might have been different if more severely affected patients were included. However, since the person separation index of the R-ODS turned out to be extremely high (0.97) it is reasonable to assume that this scale will also be able to differentiate between the various functional phase changes in patients with GBS.[22]

The AMC linear disability score (ALDS) item bank and ACTIVLIM are both linearly weighted scales quantifying functional status. The ALDS focuses on chronic conditions and ACTIVLIM on various neuromuscular disorders. We compared the item locations of the R-ODS with the locations of similar items of the ALDS and the ACTIVLIM.[13, 15] Not only do the person locations (logits) differ considerably between these scales and the R-ODS, but also the item difficulty ordering differs noticeably. This strengthens the idea that disease-specific scales should be constructed. In our view, although activity and social participation limitations are considered not being disease-specific, the significance and weight of corresponding items may appear

specific for the various illnesses. Support for this implication is also obtained when looking at the responses of patients having multifocal motor neuropathy (MMN).[30] Our primary aim was to construct an outcome measure suitable for use in future trials evaluating therapeutic efficacy in a wide range of immune-mediated neuropathies. Therefore, at first not only the current patient population (GBS, CIDP and MGUSP), but also patients with stable MMN were included. However, the primary analyses showed the MMN patients behaving strongly differently. Fine motor upper limb items like 'turn a key in a lock' and 'make a sandwich' were considered much more difficult to accomplish by the MMN group (thus showing significant uniform DIF on personal factor 'diagnosis'). Their responses to the above items resembled the difficulty levels of items like 'walking uphill' and 'standing for hours' in our patients with GBS, CIDP and MGUSP. We decided to not split the deviating items for the various illnesses. Instead, we are currently constructing a MMN-specific activity and participation scale to bypass these limitations (see also chapter 3.4).

Both item response theory (IRT) and Rasch measurement estimate the probability that a person with a given ability level will be able to complete an item/task.[9, 11, 31] The ALDS is based on IRT, which aims to find the best model that explains the data whereas data based on Rasch have to satisfy the model expectations.[32-34] As a result, in IRT often a model with generally fewer restrictions is chosen to explain all data. We developed the R-ODS using the Rasch model since it is considered a strong model with many restrictions aiming to satisfy mathematical requirements necessary to achieve fundamental measurement.[35]

There are some limitations that should be addressed. Most participants (86%) were members of a patient organisation, the Dutch society of neuromuscular disorders (Vereniging Spierziekten Nederland [VSN]). We could not verify their diagnosis, but the board of the VSN has ascertained us that all patients came from centres with great expertise in neuromuscular disorders, ensuring the right diagnosis. Written instructions were given how to respond when assistance or special devices were needed. Nevertheless, there might be some bias of the responses given, depending on the daily and social situation of patients and adaptations made. Furthermore, due to cultural and geographic differences items may be applicable to the Netherlands but not necessarily elsewhere. The ability of the R-ODS to detect relevant clinical changes over time (responsiveness) also needs further evaluation. Its responsiveness and international applicability is currently being investigated in patients with newly diagnosed immune-mediated neuropathies as part of the international multi-centre *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study.

Nevertheless, the R-ODS seems to be a valid and reliable outcome measure capturing activity and social participation limitations in patients with GBS, CIDP and MGUSP. Compared to the ODSS, the R-ODS represents a wider range of item

difficulties, thereby showing a better targeting of the different ability levels of these patients. If its responsiveness can be demonstrated, we expect that the R-ODS will be valuable for future clinical trials and follow-up studies in patients with these disorders.

References

1. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
2. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
3. Hughes, R.A., et al., *Controlled trial prednisolone in acute polyneuropathy.* Lancet, 1978. **2**(8093): p. 750-3.
4. van Swieten, J.C., et al., *Interobserver agreement for the assessment of handicap in stroke patients.* Stroke, 1988. **19**(5): p. 604-7.
5. Merkies, I.S., et al., *Clinimetric evaluation of a new overall disability scale in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2002. **72**(5): p. 596-601.
6. Merkies, I.S., et al., *Psychometric evaluation of a new handicap scale in immune-mediated polyneuropathies.* Muscle Nerve, 2002. **25**(3): p. 370-7.
7. Graham, R.C. and R.A. Hughes, *A modified peripheral neuropathy scale: the Overall Neuropathy Limitations Scale.* J Neurol Neurosurg Psychiatry, 2006. **77**(8): p. 973-6.
8. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.
9. Rasch, G., *Probabilistic models for some intelligence and attainment tests.* 1960, Copenhagen: Danmarks Paedagogiske Institut.
10. Tennant, A. and P.G. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* Arthritis Rheum, 2007. **57**(8): p. 1358-62.
11. Hambleton, R.K., H. Swaminathan, and H.J. Rogers, *Fundamentals of item response theory*. 1991, London: Sage.
12. Streiner, D.L. and G.R. Norman, *Health measurement scales. A practical guide to their development and use.* 2nd ed. 1998, New York: Oxford University Press.
13. Holman, R., et al., *The AMC Linear Disability Score project in a population requiring residential care: psychometric properties.* Health Qual Life Outcomes, 2004. **2**: p. 42.
14. Holman, R., et al., *The Academic Medical Center Linear Disability Score (ALDS) item bank: item response theory analysis in a mixed patient population.* Health Qual Life Outcomes, 2005. **3**: p. 83.
15. Vandervelde, L., et al., *ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders.* Neuromuscul Disord, 2007. **17**(6): p. 459-69.
16. Hobart, J.C., D.L. Lamping, and A.J. Thompson, *Evaluating neurological outcome measures: the bare essentials.* J Neurol Neurosurg Psychiatry, 1996. **60**(2): p. 127-30.
17. Bombardier, C. and P. Tugwell, *A methodological framework to develop and select indices for clinical trials: statistical and judgmental approaches.* J Rheumatol, 1982. **9**(5): p. 753-7.
18. World Health Organization, *International classification of impairments, disabilities, and handicaps*. 2001: Geneva.
19. Andrich, D., et al., *Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0)*. 2003, Rumm Laboratory Pty Ltd.: Duncraig, Western Australia.
20. Hermans, M.C., et al., *Rasch-built myotonic dystrophy type 1 activity and participation scale (DM1-Activ)*. Neuromuscul Disord, 2010. **20**(5): p. 310-8.
21. Linacre, J., *Sample size and item calibration stability.* Rasch Meas Trans, 1994. **7**: p. 328.
22. Fischer, W.P., *Reliability statistics.* Rasch Meas Trans, 1992. **6**: p. 238.
23. Wright, B.D. and M.H. Stone, *Best test design*. 1979, Chicago: Mesa Press.
24. Bland, J.M. and D.G. Altman, *Multiple significance tests: the Bonferroni method.* Bmj, 1995. **310**(6973): p. 170.
25. Comi, G., et al., *A randomised controlled trial of intravenous immunoglobulin in IgM paraprotein associated demyelinating neuropathy.* J Neurol, 2002. **249**(10): p. 1370-7.
26. Hughes, R.A., et al., *Intravenous immune globulin (10% caprylate-chromatography purified) for the treatment of chronic inflammatory demyelinating polyradiculoneuropathy (ICE study): a randomised placebo-controlled trial.* Lancet Neurol, 2008. **7**(2): p. 136-44.
27. RMC trial group, *Randomised controlled trial of methotrexate for chronic inflammatory demyelinating polyradiculoneuropathy (RMC trial): a pilot, multicentre study.* Lancet Neurol, 2009. **8**(2): p. 158-64.
28. van Koningsveld, R., et al., *Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barre syndrome: randomised trial.* Lancet, 2004. **363**(9404): p. 192-6.
29. Pritchard, J., et al., *A randomized controlled trial of recombinant interferon-beta 1a in Guillain-Barre syndrome.* Neurology, 2003. **61**(9): p. 1282-4.
30. Van Asseldonk, J.T., et al., *Multifocal motor neuropathy.* Lancet Neurol, 2005. **4**(5): p. 309-19.
31. Hays, R.D., L.S. Morales, and S.P. Reise, *Item response theory and health outcomes measurement in the 21st century.* Med Care, 2000. **38**(9 Suppl): p. II28-42.
32. Massof, R.W., *The measurement of vision disability.* Optom Vis Sci, 2002. **79**(8): p. 516-52.
33. Hobart, J. and S. Cano, *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.* Health Technol Assess, 2009. **13**(12): p. iii, ix-x, 1-177.
34. Tesio, L., *Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research.* J Rehabil Med, 2003. **35**(3): p. 105-15.
35. Andrich, D., *Controversy and the Rasch model: a characteristic of incompatible paradigms?* Med Care, 2004. **42**(1 Suppl): p. I7-16.

Chapter 3.4

Rasch-built overall disability scale for multifocal motor neuropathy (R-ODS-MMN)

Els K. Vanhoutte[1], Sonja I. van Nes[2], Elisabeth A. Cats[3], Willem L. van der Pol[3], Kenneth C. Gorson[4], Francesca Gallia[5], Francesco Bombelli[6], Richard Lewis[7], Pieter A. van Doorn[2], Peter Y.K. Van den Bergh[8], Jean-Marc Léger[6], Eduardo Nobile-Orazio[5], David R. Cornblath [9], Leonard H. van den Berg[3], Ingemar S.J. Merkies[1,2,10], C.G. Faber[1], on behalf of the PeriNomS study group

[1]Department of Neurology, University Medical Centre Maastricht, Maastricht, the Netherlands
[2]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[3]Department of Neurology, Rudolf Magnus Institute of Neuroscience University Medical Centre, Utrecht, the Netherlands
[4]Department of Neurology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, USA
[5]Department of Neurological Sciences, Milan University, Humanitas Clinical Institute, Rozzano, Milan, Italy
[6]Department of Neurology, Hopital de la Salpêtrière, Paris, France
[7]Department of Neurology, Wayne State University School of Medicine, Detroit, USA
[8]Department of Neurology, Catholique University of Louvain, Belgium
[9]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, USA
[10]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands

Submitted for publication

Abstract

Background: Disability outcome measures that have been used in studies of patients with multifocal motor neuropathy (MMN) are ordinal based and may not be ideally suited to accurately capture functional changes after treatment. Currently, there is no MMN-specific scale with a linear construct as an alternative.

Objective: To construct an interval outcome measure suitable to capture the limitations in daily and social activities of patients with MMN using the Rasch model.

Methods: 146 preliminary activity and participation items were assessed twice (interval: 2-4 weeks; test-retest reliability studies) in 107 MMN patients and subjected to Rasch analyses. The ordinal based overall disability sumscore (ODSS) was also assessed for validity purposes. The final Rasch-built overall disability scale for MMN (R-ODS-MMN) was longitudinally applied (period: 1 year) in 7 patients with newly diagnosed MMN receiving intravenous immunoglobulin to capture its responsiveness.

Results: 121 items were removed step by step due to insufficient face and content validity, disordered thresholds, misfit statistics, item bias, and local dependency. The 25 remaining items were ordered on a linearly weighted scale and demonstrated good validity, reliability, and responsiveness.

Conclusion: The 25 item-R-ODS-MMN is a disease-specific, interval measure suitable to detect activity and social participation limitations in patients with MMN, overcoming the shortcomings of previous scales. Its use is recommended for future clinical trials in patients with MMN.

Introduction

Multifocal motor neuropathy (MMN) is an uncommon immune-mediated demyelinating disorder characterised by slowly progressive, predominantly distal, asymmetrical limb weakness with involvement of motor nerves. The diagnosis is mainly based on clinical and electrophysiological findings.[1] Patients with MMN benefit from intravenous immunoglobulin (IVIg) therapy based upon randomised cross-over trials using the Medical Research Council grading system as the primary impairment outcome measure.[1-7] From a review of these trials, no significant difference was seen at the disability level.[8] Most studies in MMN have used the (modified) Rankin scale to assess disability.[3, 4, 6, 7, 9, 10] The Guy's neurological disability scale, the nine-hole pegboard test, a self-evaluation scale, and the Norris scale are examples of other disability scales or composite scales containing disability items that have been used in MMN.[3, 4, 11-19] Despite having deficiencies from modern clinimetric perspective, these disability outcome measures have been generally used in MMN in the absence of linearly constructed ability and participation measures.[20, 21] Also, these measures may be less sensitive to capture relevant clinical changes over time and therefore may have contributed to the non-significant difference between patients receiving IVIg compared to placebo.[8]

The aim of this paper is to present the construction of the Rasch-built overall disability scale specifically designed to capture the limitations in daily and social activities of patients with MMN (R-ODS-MMN).

Methods

Patients

Cross-sectional group

A total of 107 patients, aged 18 years and older with a stable clinical condition (of which 68 have been receiving maintenance interval therapy with IVIg) were recruited between July 2009 and May 2010. All patients met the clinical and electrophysiological diagnostic criteria for MMN.[22] A clinically stable condition was defined as unchanged activities in daily living as declared by the patient, or no objective changes at neurological examination when compared with previous clinical findings ≥ two months before the start of the study. Eighty-three patients were recruited from the neuromuscular research group at the Rudolf Magnus institute of neuroscience, University Medical Centre Utrecht, the Netherlands, and 24 were recruited from other centres with expertise in MMN: 10 from Italy, 9 from France, 4 from the USA, and 1 from Belgium.

Longitudinal group

Seven newly diagnosed patients with MMN were enrolled to investigate the responsiveness of the newly constructed disability scale. These patients were all

treated with IVIg.

For both the cross-sectional and the longitudinal group, patients could not be included if concomitant diseases (e.g., diabetes, renal insufficiency, (prior) chemotherapy, alcohol abuse (> 5 IU/day)) could interfere with general nervous system as well as physical functioning.

Standard protocol approvals, registrations, and patient consents
The local medical ethics committee in each participating centre approved the protocol. Written informed consent was obtained from all participants.

Questionnaire development
Accepted standardised scale development procedures were applied to create the MMN-specific activity and participation scale, similar to the construction of the Rasch-built overall disability scale (R-ODS) for patients with Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyneuropathy (CIDP), and polyneuropathy associated with monoclonal gammopathy of undetermined significance (MGUSP).[23-25] In brief, we critically reviewed potential items at the activity and participation level using the WHO International Classification of Functioning, Disability and Health (ICF), systematic literature review with reference tracing, and interviews with a focus group of 12 MMN patients.[26] Information elicited from this group was subsequently combined with earlier selected items to form a large item pool, the so-called preliminary MMN-specific Rasch-built overall disability scale (preliminary R-ODS-MMN), containing 146 items. Patients were asked to rate their perceived difficulty to perform the selected items on a three-level scale: 0 ('unable to perform'), 1 ('able to perform, but with difficulty'), and 2 ('easily performed, without difficulty'). An item was scored 3 if it was not applicable to the patient.

Additional outcome measure
The overall disability sumscore (ODSS) is a previously validated scale in patients with GBS and CIDP that comprises a functional description of the arms and legs and ranges from 0 ('no signs of disability') to 12 ('most severe disability score'; arm grade: 0–5, leg grade 0-7).[27] The ODSS was applied in the cross-sectional group of patients to examine construct validity of the final R-ODS-MMN.

Procedures
After presenting standardised instructions, the cross-sectional group was requested to complete the preliminary R-ODS-MMN questionnaire twice (test-retest study; interval 2-4 weeks; 1st assessment: n=107 patients, 2nd assessment: n=77 patients returned the questionnaire). The ODSS was assessed once in the cross-sectional group (validity study). The longitudinal group completed the final R-ODS-MMN at entry, 3, and 12 months of follow-up.

Rasch analyses and statistical aspects
The preliminary R-ODS-MMN was subjected to Rasch analysis to determine whether model expectations were met. This methodology has been described thoroughly elsewhere, also specifically for neurologists (see also chapter 3.1).[25, 28-30] In brief, this statistical technique transforms ordinal obtained scores (which are scale dependent and of limited accuracy), into interval measures that are scale independent and suitably accurate for individual patient assessment. This method is based on the logical assumption that individuals with greater ability to perform a particular task should have an increased probability, relative to individuals with lower ability levels, of achieving a higher score on the item in question.[28, 30] Analyses were performed to obtain a final R-ODS-MMN meeting all Rasch expectations, such as proper fit statistical parameters, lack of item bias, no disordered thresholds or local dependency, and fulfilment of unidimensionality.[25, 28, 30]

Descriptive statistics and sample size
Personal factors like age, gender, duration of symptoms, and country of assessment were collected. For the purposes of the current study, these person factors were categorised as follows: age (<40 years vs. 40-49 years vs. 50-59 years vs. ≥60 years), gender (female vs. male), duration of complaints (<5 years vs. 5-9 years vs. 10-20 years vs. ≥20 years), and country (Holland/Belgium vs. France vs. Italy vs. USA). Age and duration factors were categorised striving for equivalent distribution of participants per subgroup. According to sample size rules, at least 150 records were needed for acceptable model stability.[32] To fulfil this requirement, the data of the first and second assessments were stacked, controlling for 'time factor' as a possible confounding factor.[33] These procedures led to a sample size of 184 patient records (1st and 2nd assessment) for examination.

Reliability and validity
Internal reliability was examined by determining the person separation index (PSI). A PSI above 0.7 is considered acceptable, indicating the ability to identify at least two groups of patients.[31] Test-retest reliability studies (patients' ability locations) were also performed to determine the consistency of the final scale created.[34] Reliability was quantified by calculation of the intraclass correlation coefficient using a one-way random effects analysis-of-variance (ANOVA) model for group' comparisons. The external construct validity of the final scale was assessed by correlation with the ODSS.

Responsiveness
Traditional responsiveness at the longitudinal *group level* was calculated using the effect size (ES) indicator at 3 and 12 months of follow-up. The ES is equal to the mean change in score divided by the standard deviation of the scores at

entry (ES=μi-μo/SD_{entry}); μi = mean final R-ODS-MMN score of the longitudinally examined group at month = *i*; μo = mean final R-ODS-MMN score at entry).[35] Responsiveness at the *individual person level* was also calculated. Modern clinimetric approaches have demonstrated that the standard error (SE) around an individual patient location may vary across the theoretical range of an outcome measure, and with this also the clinical importance of changes within a patient over time.[36, 37] Minimal clinically important difference-standard error (MCID-SE), as a score for individual responsiveness, was calculated based on the previously described significant change (SigChange).[38] In brief, MCID-SE was calculated by computing for each of the seven serially examined patients separately: a) their own change (person location at *i* month minus person location at entry; where *i* = 3 or 12 months of follow-up), b) the corresponding SE of difference related to their individual change (SE_{diff} = square-root(SE_{entry}^2+$SE_{i\ month}^2$), and c) the final MCID-SE calculations by dividing the individual change scores by corresponding SE_{diff} (MCID-SE = (person location at *i* month minus person location at entry)/SE_{diff}). Since MMN is considered an indolent illness, we arbitrarily defined the minimum clinically important difference cut-off using ≥ 1 standard error (corresponding to 68% certainty that the ability estimate is no measurement error). As a result, the following subgroups were defined:

- subgroup 1 (clinically important improvement): MCID-SE ≥ 1
- subgroup 2 (clinically unimportant improvement): 0 < MCID-SE < 1
- subgroup 3 (no change): MCID-SE = 0
- subgroup 4 (clinically unimportant worsening): -1 < MCID-SE < 0
- subgroup 5 (clinically important worsening): MCID-SE ≤ -1.

Software

Rasch analyses were performed using Rasch Unidimensional Measurement Models (RUMM2030), with the partial credit model as default. Further analyses were undertaken using Stata 11.0 for Windows XP. The p-value was adjusted (Bonferroni) throughout the analyses correcting for multiple testing.[39]

Results

Study population and data quality control

The study population demographics are presented in table 1. Most patients were men. Based on the clinical characteristics of patients with MMN, a total of 19 items had insufficient face and content validity and were removed. In the model construction, items scored as 3 ('not applicable') were interpreted as missing data. In addition, a total of 16 items with > 10% missing values were omitted as part of the quality control procedure. A total of 111 items were kept and subjected to Rasch analyses.

Table 1. General characteristics of patients with multifocal motor neuropathy

	Cross-sectional group (for R-ODS-MMN scale construction including validity and reliability studies)	Longitudinal group (for R-ODS-MMN responsiveness studies)
Number of patients	107	7
Age (years); mean (SD), range	52.58 (11.4), 29 – 86	49.4 (9.3), 35 - 61
Gender; n (%) female male	 25 (23.4) 82 (76.6)	 1 (14.3) 6 (85.7)
Duration of symptoms (years); mean (SD), range	12.8 (7.8), 0.2 – 46	8 (3.7), 3 - 12
Country of assessment; n (%) Holland Italy France USA Belgium	 83 (77.6) 10 (9.4) 9 (8.4) 4 (3.7) 1 (0.9)	 2 (28.6) 1 (14.3) 4 (57.1)
ODSS arm grade; (%) ≤ 1 (normal or minor symptoms, not affecting any) 2 (moderate symptoms, affecting but not preventing any) 3 (severe symptoms, preventing at least one but not all) ≥ 4 (severe symptoms, preventing all or no purposeful movements)	 19.1% 56.2% 22.5% 2.2%	
ODSS leg grade; (%) 0 – 1 (normal or affected walking, but looks normal) 2 – 4 (walking looks abnormal; unilateral/bilateral support) 5 – 7 (wheelchair or bedbound)	 40.5% 56.2% 3.3%	-

Initial Rasch analyses on the preliminary R-ODS-MMN

The remaining 111 items of the preliminary R-ODS-MMN showed overall misfit. The items fit residuals showed acceptable fit statistics (mean: -0.435; SD: 1.196), whereas the obtained scores for the person fit residuals deviated from model expectations (mean: -0.455, SD: 1.523). The significant chi-square probability ($p<0.00001$) demonstrated no invariance of items and a proportion of 0.25 of the t-tests performed fell outside the ± 1.96 range, indicating multidimensionality.

Data handling to fit the Rasch model

Throughout the analyses, we continuously monitored the distribution of persons within the class intervals. In order to improve the model, items and persons not fulfilling the Rasch model requirements were step by step evaluated and removed one by one if needed.

Step 1: Two items ('get in a car' and 'drive a car') demonstrated disordered thresholds and were removed.

Step 2: The individual item fit statistics of 10 items demonstrated misfit to the model (having a significant chi-square probability or having fit residuals exceeding ± 2.5) and were removed.

Step 3: Seven items demonstrated item bias: 5 items had differential item functioning (DIF) on personal factor 'country' (2 uniform, 2 non-uniform, 1 both), and 2 items demonstrated DIF (1 uniform, 1 non-uniform) on personal factor 'duration'. We systematically removed these 7 items. Also, 4 cross-sectional patients demonstrated DIF (using the 95th confidence intervals) on time factor and were removed as well.

Step 4: Numerous local dependency findings were found between many items. All item sets with correlations above 0.28 were evaluated starting with the highest correlations (>0.7, subsequently >0.6, etc., to >0.28). Of each item set, the item showing less clinical relevance (face and content validity according to two experts having consensus on their judgment) and with most over- or under-discrimination on its category probability curve was removed. A total of 67 items were stepwise removed.

After completing these procedures, we succeeded in constructing a 25-item interval measure (R-ODS-MMN) that met all Rasch model expectations (item fit residuals: mean -0.352, SD 0.997; person fit residuals: mean -0.333, SD 0.982; item-trait chi-square: p-value=0.10, DF=50). Based on the first principal components analysis two subsets of items were formed (6 most positively loading vs. 6 most negatively loading items). The independent t-tests between these two subsets suggested acceptable unidimensionality (0.052 (95% confidence interval (CI): 0.019-0.084). In the final R-ODS-MMN scale, the item 'read a newspaper' was the easiest to perform while 'serve coffee/tea on a tray' turned out to be the most difficult task (table 2). Item difficulty ranged from -2.877 to 2.806 logits. A total of 18 patients

demonstrated a maximum score (9.8%). The raw scores were translated to an interval measure with logits as unit. These less intuitive logits may be translated to a more understandable centile metric ranging from 0 (most severe activity and participation restrictions) to 100 (no activity limitations and participation restrictions) (nomogram available on request).

Table 2. Final 25-item Rasch-built overall disability scale for multifocal motor neuropathy (R-ODS-MMN)

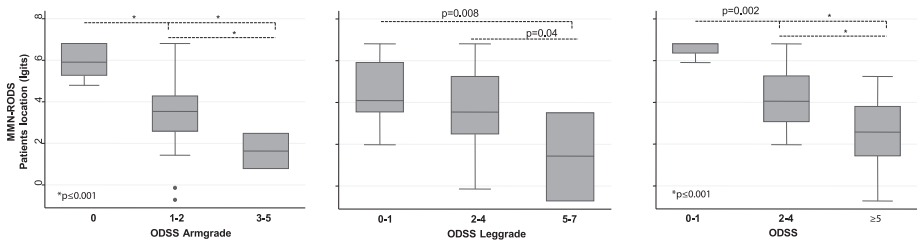
	Are you able to:	unable to perform	able to perform, but with difficulty	able to perform without difficulty
		0	1	2
1	Read a newspaper?			
2	Eat?			
3	Make a telephone call?			
4	Open or close a door?			
5	Brush your teeth?			
6	Drink out of a glass?			
7	Turn a key in a lock?			
8	Use knife and fork (spoon)?			
9	Wash upper body?			
10	Wash lower body?			
11	Dress upper body?			
12	Shave/epilate?			
13	Clean after toilet visit?			
14	Zip trousers?			
15	Fill in a form?			
16	Get money from a cash point?			
17	Work on a computer?			
18	Vacuum cleaning?			
19	Dress lower body?			
20	Catch an object (e.g., ball)?			
21	Slice vegetables?			
22	Handle small objects (e.g., coin)?			
23	Tie laces?			
24	Clip nails?			
25	Serve coffee/tea on a tray?			

Legend to table 2. The easiest item turned out to be 'read a newspaper', the most difficult item 'serve coffee/tea on a tray'. The raw scores can be translated to the less intuitive logits, and from here to a more understandable centile metric ranging from 0 (most severe activity and participation restrictions) to 100 (no activity limitations and participation restrictions).

Validity and reliability studies

The final 25-item R-ODS-MMN scores demonstrated acceptable construct validity when correlated with the ODSS scores (figure 1).

Figure 1. Construct validity demonstrated for the 25-item R-ODS-MMN when compared to the scores of the overall disability sumscore (ODSS)



Legend to figure 1. A high patient location on the R-ODS-MMN refers to minor activity and participation limitations. Low scores on the ODSS also refer to less disability. This figure visualises the correlations between R-ODS-MMN patient locations and ODSS scores, reflecting acceptable construct validity.

Internal reliability remained robust as indicated by a person separation index of 0.87. Also, the test-retest reliability for personal location was good: patients' locations were almost always located within the 95% CI lines, reflecting ideal invariance (R^2=0.87).

Responsiveness studies

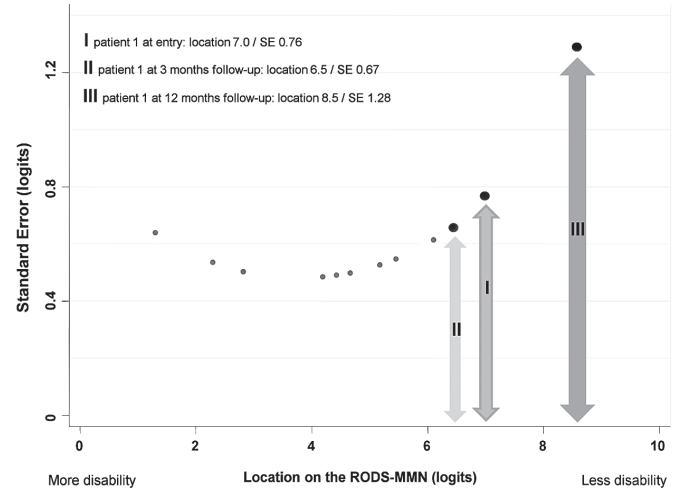
At 3 months, only 1 patient had an improvement in R-ODS-MMN score; 3 patients did not show any changes, while 3 others demonstrated some deterioration. At 12 months, 6 patients had higher scores for the R-ODS-MMN and 1 patient had unchanged values. The effect size values were poor (at 3 and 12 months 0.1 and 0.4, respectively). However, responsiveness at the *individual person level,* taking the changing standard errors into account, demonstrated a different dynamic (table 3 and figure 2): at 3 months, 1 patient (#4) demonstrated a clinically important deterioration compared to entry (subgroup 5), while 2 others (#1 and #6) deteriorated although not clinically important (subgroup 4), 3 patients (#2,#3 and #7) remained unchanged (subgroup 3), and 1 (#5) showed improvement although clinically not important (table 3). At 12 months, 2 patients (#1 and #3) demonstrated a clinically important improvement compared to entry (subgroup 1), 4 others (#2, #4-6) improved clinically not important (subgroup 2; having MCID-SE scores: 0.71-0.89), and 1 patient (#7) remained unchanged (table 3).

Table 3. Categorising MCID-SE scores of each longitudinally examined patient (n=7)

Patient #	MCID-SE at 3 months	MCID-SE at 12 months	subgroup classification at 3 months compared to entry	subgroup classification at 12 months compared to entry
1	-0.47	1.04	4	1
2	0	0.89	3	2
3	0	1.73	3	1
4	-1.2	0.71	5	2
5	0.79	0.79	2	2
6	-0.34	0.71	4	2
7	0	0	3	3

Legend to table 3. MCID-SE scores were determined by dividing the individual change scores by corresponding SE_{diff} (MCID-SE = (person location at *i* month minus person location at entry)/SE_{diff} ; where *i* = 3 or 12 months of follow-up and SE_{diff}= square-root(SE_{entry}^2+$SE_{i\ month}^2$)).
Subgroup classification: 1 (clinically important improvement: MCID-SE ≥ 1), 2 (clinically unimportant improvement: 0 < MCID-SE < 1), 3 (no change: MCID-SE = 0), 4 (clinically unimportant worsening: -1 < MCID-SE < 0), 5 (clinically important worsening: MCID-SE ≤ -1).

Figure 2. Patient location on the R-ODS-MMN with corresponding standard error



Legend to figure 2. An 'U'-shape pattern is shown indicating that the standard error changes with the changing location of a patient on the theoretical range of the R-ODS-MMN. As an example: the black dots correspond with the location of patient 1 at entry (I), at 3 months (II), and 12 months (III) of follow-up. Note the changing corresponding standard error with each time point. See also table 3 for the categorisation of changes seen in these patients.

Discussion

The R-ODS-MMN scale is a linearly weighted outcome measure constructed specifically to capture activity and participation limitations in patients with MMN. This scale was developed noting the limitations seen in most disability ordinal based outcome measures used thus far in MMN.[20] The R-ODS-MMN fulfilled all Rasch model expectations and demonstrated acceptable validity, reliability, and responsiveness. These findings contribute to the paradigm shift needed to improve monitoring of treatment and follow-up of patients with chronic conditions like MMN.

The clinical presentation of a disease determines which items or tasks will be completed easily or with more effort. For example, based on differences in the clinical features, it is logical to expect that patients with MMN will experience more difficulty with fine motor upper limb items compared to patients with GBS. We previously created a R-ODS scale for patients with GBS, CIDP and MGUSP.[25] Comparing the R-ODS for GBS/CIDP/MGUSP vs. the R-ODS-MMN, for similar and beforehand non-disease specific items used for their construction, the locations (weights or difficulties) of items included in the final scales are certainly disease specific and varied considerably. For example, for patients with MMN, the item 'serve coffee/tea on a tray' (location: 2.806 logits) was the most difficult to accomplish, whereas in GBS/CIDP/MGUSP patients the 'ability to run' (location: 5.604 logits) turned out to be the most difficult to perform. For GBS/CIDP/MGUSP patients the item 'wash upper body' was relatively easy to accomplish (location: -2,5716), but for patients with MMN this item was considerably more difficult to accomplish (location -0,320).[25]

Other outcome measures with a linear construct capturing disability in neuromuscular disorders are the ACTIVLIM and the AMC linear disability score (ALDS).[40, 41] However, these scales should be considered as generic based measures for chronic illnesses with the ACTIVLIM focusing on numerous neuromuscular disorders. Furthermore, little common ground was found in the difficulty of the items between these scales and the R-ODS-MMN since only 3 items demonstrated the same content. In MMN distal arm muscles are mainly affected, it is therefore not surprising that most items address distal muscle functioning. However, the ACTIVLIM focuses on neuromuscular disorders in general and this is reflected by the content of the selected items.

The effect sizes of the longitudinally examined patients were poor at 3 and 12 months of follow-up. However, using traditional responsiveness indicators like the effect size does not always provide information on the personal magnitude and direction (improvement, stable, or deterioration) of changes for each individual patient as is the case when using modern tools like the Rasch method. Hence, traditional methods tend to be misleading and do not take the changing measurement errors observed in individual patients into consideration (figure 2).[28-30, 36, 37] The clinical dynamics were clearly captured by the Rasch method, showing improvement at 12 months in 6 of the 7 patients, although this was only clinically important in 2 patients (table 3).

Some methodological limitations in this study need to be addressed. In the ideal situation, a sample size of 150, but preferably approximately 250 patients is needed to provide accurate model stability.[32] The minimal requirement could only be achieved after stacking the data to a total sample size of 184 records. However, for an orphan disease like MMN this is a large series of patients. Further efforts are needed in collaboration with neuromuscular centres worldwide to provide a more stable scale for this rare disease. Second, there was a ceiling effect in 9.8%, which could hamper the applicability of the scale. Efforts have been made to extend the final R-ODS-MMN to ~30 items, but the ceiling effect remained unchanged. A proportion of patients in the cross-sectional group hardly experienced any disability related to their illness. Finally, more longitudinally examined patients with MMN are needed to assess the responsiveness of the R-ODS-MMN at the individual person level. Efforts for this are currently underway as part of the *P*eripheral *N*europathy outcome measures standardisation (***P*eri*N*om*S***) study, an international multi-centre study to improve the clinical assessment in patients with inflammatory neuropathies. Despite these limitations, the R-ODS-MMN fulfils all Rasch model requirements and therefore substantially increases our ability to accurately measure activity and participation limitations to capture clinically important changes in patients with MMN.

References

1. Van Asseldonk, J.T., et al., *Multifocal motor neuropathy.* Lancet Neurol, 2005. **4**(5): p. 309-19.
2. Medical Research Council, *Aids to the investigation of the peripheral nervous system.* 1943, London: Her Majesty's Stationanery Office.
3. Meucci, N., et al., *Long term effect of intravenous immunoglobulins and oral cyclophosphamide in multifocal motor neuropathy.* J Neurol Neurosurg Psychiatry, 1997. **63**(6): p. 765-9.
4. Nobile-Orazio, E., et al., *Multifocal motor neuropathy: clinical and immunological features and response to IVIg in relation to the presence and degree of motor conduction block.* J Neurol Neurosurg Psychiatry, 2002. **72**(6): p. 761-6.
5. Umapathi, T., et al., *Immunosuppressant and immunomodulatory treatments for multifocal motor neuropathy.* Cochrane Database Syst Rev, 2009(1): p. CD003217.
6. Van den Berg, L.H., et al., *Treatment of multifocal motor neuropathy with high dose intravenous immunoglobulins: a double blind, placebo controlled study.* J Neurol Neurosurg Psychiatry, 1995. **59**(3): p. 248-52.
7. Van den Berg-Vos, R.M., et al., *Treatment of multifocal motor neuropathy with interferon-beta1A.* Neurology, 2000. **54**(7): p. 1518-21.
8. van Schaik, I.N., et al., *Intravenous immunoglobulin for multifocal motor neuropathy.* Cochrane Database Syst Rev, 2005(2): p. CD004429.
9. Rankin, J., *Cerebral vascular accidents in patients over the age of 60: prognosis.* Scott Med J, 1957. **2**: p. 200-15
10. UK-TIA study group. *The UK-TIA aspirin trial: interim results.* Br Med J, 1988. **296**: p. 316-20
11. Sharrack, B. and R.A. Hughes, *The Guy's neurological disability scale (GNDS): a new disability measure for multiple sclerosis.* Mult Scler 1999. **5** (4): p. 223-33
12. Oxford, G.K. et al., *Adult norms for a commercially available nine-hole peg test for finger dexterity.* Am J Occup Ther, 2003. **57**: p. 570-3
13. Leger, J.M., et al., *Intravenous immunoglobulin therapy in multifocal motor neuropathy: a double-blind, placebo-controlled study.* Brain, 2001. **124**(Pt 1): p. 145-53.
14. Norris, F.H. Jr., et al., *The administration of guanidine in amyotrophic lateral sclerosis.* Neurology, 1974. **24**: p. 721-28
15. Azulay, J.P., et al., *Intravenous immunoglobulin treatment in patients with motor neuron syndromes associated with anti-GM1 antibodies: a double-blind, placebo-controlled study.* Neurology, 1994. **44**(3 Pt 1): p. 429-32.
16. Baumann, A., C.W. Hess, and M. Sturzenegger, *IVIg dose increase in multifocal motor neuropathy: a prospective six month follow-up.* J Neurol, 2009. **256**(4): p. 608-14.
17. Cats, E.A., et al., *New liquid intravenous immunoglobulin (10 % IVIg) for treatment of multifocal motor neuropathy: a prospective study of efficacy, safety and tolerability.* J Neurol, 2008. **255**(10): p. 1598-9.
18. Piepers, S., et al., *Mycophenolate mofetil as adjunctive therapy for MMN patients: a randomised, controlled trial.* Brain, 2007. **130**(Pt 8): p. 2004-10.
19. van Nes, S.I., C.G. Faber, and I.S. Merkies, *Outcome measures in immune-mediated neuropathies: the need to standardize their use and to understand the clinimetric essentials.* J Peripher Nerv Syst, 2008. **13**(2): p. 136-47.
20. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
21. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.
22. van Schaik, I.N., et al., *European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of multifocal motor neuropathy.* Eur J Neurol, 2006. **13**(8): p. 802-8.
23. Streiner, D.L. and G.R. Norman, *Health measurement scales. A practical guide to their development and use.* 2nd ed. 1998, New York: Oxford University Press.
24. Bombardier, C. and P. Tugwell, *A methodological framework to develop and select indices for clinical trials: statistical and judgmental approaches.* J Rheumatol, 1982. **9**(5): p. 753-7.
25. van Nes, S.I., et al., *Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies.* Neurology, 2011. **76**(4): p. 337-45.

26. *W.H.O. The International Classification of Functioning, Disability and Health (ICF)*. 2010, WHO: Geneve.
27. Merkies, I.S., et al., *Clinimetric evaluation of a new overall disability scale in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2002. **72**(5): p. 596-601.
28. Rasch, G., *Probabilistic models for some intelligence and attainment tests.* 1960, Copenhagen: Danmarks Paedagogiske Institut.
29. Andrich, D., et al., *Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0)*. 2003, Rumm Laboratory Pty Ltd.: Duncraig, Western Australia.
30. Tennant, A. and P.G. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* Arthritis Rheum, 2007. **57**(8): p. 1358-62.
31. Fisher, W.P., *Reliability statistics.* Rasch Meas Transact, 1992. **6**: p. 238.
32. Linacre, J.M., *Sample Size and Item Calibration Stability.* Rasch Measurement Transactions, 1994. **7**(328).
33. Wright, B.D., *Rack and Stack: Time 1 vs. Time 2.* Rasch Measurement Transactions, 2003. **17**: p. 905-906.
34. Wright, B.D. and M.H. Stone, *Best test design*. 1979, Chicago: Media Press.
35. Kazis, L.E., J.J. Anderson, and R.F. Meenan, *Effect sizes for interpreting changes in health status.* Med Care, 1989. **27**(3 Suppl): p. S178-89.
36. Hobart, J.C., S.J. Cano, and A.J. Thompson, *Effect sizes can be misleading: is it time to change the way we measure change?* J Neurol Neurosurg Psychiatry, 2010. **81**(9): p. 1044-8.
37. Lai, J.-S. and D.T. Eton, *Clinically meaningful gaps.* Rasch Meas Transact, 2002. **15**: p. 850.
38. Hobart, J. and S. Cano, *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.* Health Technol Assess, 2009. **13**(12): p. iii, ix-x, 1-177.
39. Bland, J.M. and D.G. Altman, *Multiple significance tests: the Bonferroni method.* Bmj, 1995. **310**(6973): p. 170.
40. Vandervelde, L., et al., *ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders.* Neuromuscul Disord, 2007. **17**(6): p. 459-69.
41. Holman, R., et al., *The Academic Medical Center Linear Disability Score (ALDS) item bank: item response theory analysis in a mixed patient population.* Health Qual Life Outcomes, 2005. **3**: p. 83.

Chapter 4

Comparative studies

Chapter 4.1

Jamar dynamometer versus Vigorimeter to assess grip strength in immune-mediated neuropathies

Sonja I. van Nes[1], Els K. Vanhoutte[2], Mayienne Bakkers[2], Krista Kuitwaard[1], Judith Drenthen[1], Kenneth C. Gorson[3], David R. Cornblath[4], Angelika F. Hahn[5], Elisabeth A. Cats[6], Jikke-Mien M.F. Niermeijer[7], Nicolette C. Notermans[6], Leonard H. van den Berg[6], Pieter A. van Doorn[1], Catharina G. Faber[2], Ingemar S.J. Merkies[1,2,8], on behalf of the PeriNomS study group

[1]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of Neurology, Maastricht University Medical Centre, Maastricht, the Netherlands
[3]Department of Neurology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, USA
[4]Department of Neurology, Johns Hopkins Hospital and School of Medicine, Baltimore, USA
[5]Department of Neurology, London Health Science Centre, London, Ontario, Canada
[6]Department of Neurology, University Medical Centre Utrecht, Utrecht, the Netherlands
[7]Department of Pediatric Neurology, Academic Medical Centre, University of Amsterdam, the Netherlands
[8]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands

Manuscript in preparation

Abstract

Background: The Jamar dynamometer and Vigorimeter have often been used to assess grip strength in immune-mediated neuropathies, but there have been no comparison studies to determine which device is superior.

Objectives: To perform a prospective clinimetric comparison (assessing validity, reliability, and responsiveness) between the Jamar dynamometer and Vigorimeter in patients with immune-mediated neuropathies, and to determine patients' preference.

Methods: Jamar and Vigorimeter grip strength values, Medical Research Council (MRC) sumscores and overall disability sumscore (ODSS) arm grades were collected in 102 cross-sectional (stable condition) and 89 longitudinal (newly diagnosed or relapse) patients with Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyneuropathy (CIDP), monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP) or multifocal motor neuropathy (MMN). Cross-sectional patients were assessed twice for validity and reliability studies (interval 2-4 weeks). Longitudinal patients were assessed 3 to 5 times during one year, each time rating their clinical condition compared to the last visit (responsiveness studies). All patients were asked whether they preferred the Jamar or the Vigorimeter using a 5-point Likert scale.

Results: The obtained values of both devices correlated moderately with the ODSS arm grades and the MRC sumscores (Jamar: r_s =-0.58 to 0.59, Vigorimeter: r_s=-0.55 to 0.58). Similar intraclass correlation coefficients (Jamar: ICC 0.996/0.956; Vigorimeter: ICC 0.946/0.977) were found. The Guyatt responsiveness ratio was 0.64 for the Jamar and 0.56 for the Vigorimeter. More patients preferred the Vigorimeter (Vigorimeter vs. Jamar; 47% vs. 27%).

Conclusion: Validity, reliability and responsiveness aspects were comparable for the Jamar dynamometer and Vigorimeter. However, more patients preferred the Vigorimeter. Therefore, we recommend using the Vigorimeter as the standard device to assess grip strength in future studies in immune-mediated neuropathies.

Introduction

Weakness and sensory deficits in patients with immune-mediated neuropathies can have a major impact on activities of daily living, social functioning and quality of life.[1] Since distal weakness is pre-dominant in these conditions, it seems logical to use grip strength to assess outcome. Grip strength has been used as a prognostic indicator of recovery and to evaluate treatment effects in patients with diseases affecting hand function.[2-5] The Jamar dynamometer and the Vigorimeter have been used most often to assess grip strength in immune-mediated neuropathy trials.[6-9] Separately, both devices have demonstrated good scientific properties.[4, 10-20] As there is no 'gold standard', a clinimetric comparison between the Jamar dynamometer and the Vigorimeter has been suggested but has never been performed in immune-mediated neuropathies.[21] If one of these devices is a superior measure of grip strength, it could be used as a standard for future studies thus increasing precision, reliability and responsiveness analyses. Furthermore, choosing one standard grip strength device would enhance the comparability of future therapeutic trial results.

Therefore, the purpose of this study was to compare the validity, reliability and responsiveness of both devices and select the superior device as the standard grip strength measure for future studies in patients with immune-mediated neuropathies.

Methods

Patients

Cross-sectional patients

Between April 2007 and January 2010, a total of 102 patients with clinically stable immune-mediated neuropathies participated in this study: 30 patients with Guillain-Barré syndrome (GBS), 30 with chronic inflammatory demyelinating polyneuropathy (CIDP), 20 with monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP) and 22 with multifocal motor neuropathy (MMN).

Longitudinal patients

Between August 2007 and December 2010, 89 immune-mediated neuropathy patients with new or relapsing symptoms and signs participated in this study. These 36 GBS, 34 CIDP, 9 MGUSP and 10 MMN patients are participants in the on-going international multi-centre *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study.[22]

Eligibility

All patients were 18 years or older and met the international criteria for their

illness.[23-26] Only MGUSP patients with serologically proven IgM anti-MAG positive antibodies were included. Patients were excluded if they had any other potential cause for polyneuropathy, family history of neuropathy, exposure to neurotoxic medication or alcohol abuse. All cross-sectional patients had residual symptoms and signs and a stable neurological and functional status within 2 months prior to enrolment (with or without interval treatment), which was confirmed by patient interview and record review. All longitudinal patients were newly diagnosed or had a relapse after at least 2 months without any treatment.

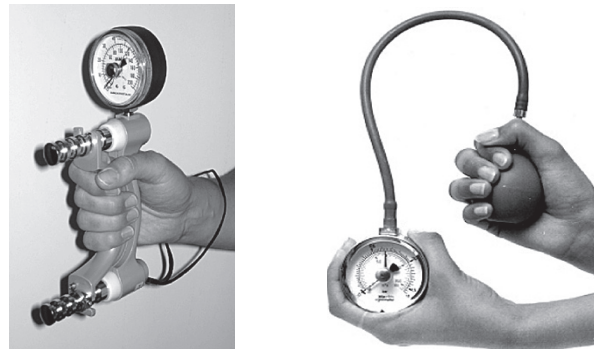
Recruitment location

Cross-sectional patients were recruited at university outpatient clinics in the Netherlands (Rotterdam, Utrecht and Maastricht). Longitudinal patients were recruited during hospital admission or at outpatient clinics in the USA (Boston and Detroit), Canada (Ontario), Brazil (Ribeirão Preto), United Kingdom (London), Italy (Milan and Venice), France (Paris), Spain (Barcelona), Belgium (Brussels) and the Netherlands (Maastricht and Rotterdam).

Outcome measures

- *Grip strength* was assessed using the Jamar hydraulic hand dynamometer (Sammons Preston Rolyan, Bolingbrook, IL, USA) and the Vigorimeter (Martin, Tuttlingen, Germany) (figure 1). The Jamar dynamometer quantifies isometric force in pounds (range 0-200). The Vigorimeter quantifies grip strength on a manometer after squeezing a rubber bulb, and is scored in kilopascal (range 0-160). All devices were newly purchased and provided to all investigators before study onset. For each patient the same set of devices was used each visit. Position two of the adjustable handle of the Jamar dynamometer was used. The middle size bulb (diameter 4.8 cm) of the Vigorimeter was used. Patients were examined sitting on a straight-backed chair, arm unsupported, feet flat on the floor and positioned according to the recommendations by the American Society of Hand Therapists; shoulder adducted and neutrally rotated, elbow flexed at 90°, forearm in neutral position, and wrist between 0°-30° dorsiflexion and between 0°-15° of ulnar deviation.[27] Three maximum voluntary contractions were recorded for each hand in alternating order (resting period: ~30 seconds), and for each trial patients were encouraged to use maximum force. The mean of the six values obtained from both hands was used for analyses. At each visit patients were asked if they had a preference for the Jamar dynamometer or the Vigorimeter using a 5-point Likert scale; 'great preference to the Jamar', 'slight preference for the Jamar', 'no preference for either device', 'slight preference for the Vigorimeter', and 'great preference for the Vigorimeter'.

Figure 1: Images of the Jamar dynamometer (A) and the Vigorimeter (B)



A. Jamar dynamometer **B.** Vigorimeter

- *The Medical Research Council (MRC) sumscore* is a summation of MRC grades given in full numbers: 0 ('no movement, no contraction'), 1 ('visible contraction without movement'), 2 ('movement but only with gravity eliminated'), 3 ('movement against gravity'), 4 ('movement against resistance, but weaker than normal'), and 5 ('normal strength') of the arm abductors, elbow flexors, wrist extensors, hip flexors, knee extensors and foot dorsal flexors on both sides, and thereby ranging from 0 (total paralysis) to 60 (normal strength).[28] An expanded MRC sumscore ranging from 0 (total paralysis) to 140 (normal strength) was used in MMN to cover its asymmetrical presentation and its often predominating weakness in distal arm and hand muscles. Based on expert opinion, the following fourteen muscle pairs, including the ten previously described as the most affected ones, were selected to compose this summation: shoulder abductors, elbow extensors and flexors, wrist extensors and flexors, finger extensors and flexors, thumb abductors, hypothenar abductors, hip flexors, knee extensors and flexors, ankle dorsiflexors and ankle plantar flexors.[29]

- *The arm grade of the overall disability sumscore (ODSS)* ranges from 0 (normal) to 5 (severe symptoms and signs in both arms preventing all purposeful movements) and can be derived from the answers to a five-question arm disability checklist ('dressing upper part of the body', 'washing/brushing hair', 'turning a key in a lock', 'using knife and fork', and 'doing/undoing buttons and zips').[30]

- *The clinical judgment score (ClinJSc)* is the score given by patients reflecting their condition. Therefore, at each visit patients were asked to answer the following question, "How would you describe your physical status now, compared to what it was on the last time you were examined?", with one of the following 5 response options: 'much worse', 'slightly worse', 'the same', 'better', and 'much better' compared to the last visit.

Study design

We standardised data collection by formally training all investigators in how to perform and record all outcome measures in a standardised manner. Also, a picture enriched training manual, demonstrating the appropriate methods of data collection, was provided to every institute participating in the ***P*eri*N*om*S*** study.

Cross-sectional patients (validity and reliability studies)

Each patient was subjected to a total of three assessments during two visits (interval 2-4 weeks). At one of two visits, two trained investigators in a blinded fashion assessed independently and consecutively grip strength using both the Jamar dynamometer and the Vigorimeter, the MRC sumscore and the arm grade of the ODSS. During the other visit one of the two investigators reassessed the same set of outcome measures without having access to the previous assessed results.

Longitudinal patients (responsiveness studies)

All patients were scheduled for examination at entry, 3 and 12 months. In patients with GBS and CIDP additional examinations at 1 and 6 months were collected to capture quicker changes in clinical status. Each visit the following assessments were performed by a trained investigator: grip strength using the Jamar dynamometer and the Vigorimeter, the MRC sumscore, the arm grade of the ODSS and the ClinJSc.

The local medical ethics committee from each participating institution approved the study protocol and all patients gave informed consent before participating.

Statistical analysis

Validity and reliability

Correlations with the ODSS arm grades and the MRC sumscores were calculated for both devices (construct validity study; Spearman's rank correlation coefficients). Our a priori hypothesis was that these correlations should be at least moderate to fulfil the requirements of convergent validity for each device. Furthermore, the intra- and inter-observer reliability scores of both devices were compared by estimating the intraclass correlation coefficients using a one-way random-effects analysis of variance (ANOVA) model. To visualise potential intra- and inter-observer measurement errors we used Bland-Altman plots.[31] The smallest detectable difference (SDD) in grip strength was computed from these plots using the formula 1.96*SD of differences. Differences in consecutive grip strength assessments greater than the SDD can be interpreted with 95% certainty as a true change (no measurement error).

Responsiveness

To determine differences in responsiveness, we correlated the difference in grip strength between all consecutive visits with the clinical judgment scores

(Spearman's rank correlation coefficients). Furthermore, for both devices the Guyatt responsiveness ratio was calculated.[32] This signal/noise ratio relates the minimum clinically important difference (MCID) to the variability in patients rating themselves as stable: MCID /standard deviation of change (SD_{change}). The MCID was defined as the mean difference in grip strength between consecutive visits of patients who rated their clinical condition as 'being better compared to the last visit' (on the ClinJSc). SD_{change} was obtained by calculating the SD of this difference in grip strength of patients rating their clinical condition as 'being stable compared to the last visit' (on the ClinJSc).

All analyses were performed using Stata 11.0 for Windows XP. A p-value (Bonferroni adjusted) of <0.05 was considered significant.[33]

Results

Descriptive statistics

Cross-sectional patients

Patient characteristics at entry are presented in table 1. Most individuals were right-handed (82%). Some patients (n=45) remained clinically stable with maintenance interval treatment. Twenty three patients with CIDP, 17 with MMN and 1 with MGUSP were being treated with periodic infusions of intravenous immunoglobulins (IVIg). Two CIDP patients were receiving plasma exchange, one CIDP patient was being treated with corticosteroids and one MGUSP patient was being treated with rituximab.

Table 1. Characteristics of the cross-sectional, clinically stable, study population (n=102) at visit one

	GBS (30)	**CIDP** (30)	**MGUSP** (20)	**MMN** (22)
Gender; male:female, n:n	13:17	21:9	17:3	17:5
Age in years; median (range)	62 (47-76)	61 (26-74)	68 (54-81)	52 (29-75)
Years since diagnosis; median (range)	11 (2-29)	6 (1-23)	9 (2-24)	12 (4-43)
Grip strength (mean both hands)				
Jamar (pounds); median (range)	63 (1-123)	61 (17-120)	81 (24-118)	22 (0-118)
Vigorimeter (kPa); median (range)	73 (17-122)	76 (18-140)	75 (23-139)	43 (0-120)
ODSS arm grade; n (%)				
0	9 (30%)	2 (7%)	3 (15%)	1 (5%)
1	5 (17%)	4 (14%)	2 (10%)	1 (5%)
2	13 (43%)	17 (58%)	10 (50%)	11 (50%)
3	3 (10%)	6 (21%)	5 (25%)	9 (41%)
MRC sumscore; median (range)	59 (48-60)	55 (46-60)	57(42-60)	109 (96-124)

Legend to table 1. kPa= kilopascal, ODSS=overall disability sumscore, MRC= Medical Research Council

Longitudinal patients

The characteristics of these 89 patients are presented in table 2. Eighty-six percent were right-handed. All patients with GBS had been treated; 23 received IVIg, 6 had plasma exchange, 5 had IVIg combined with corticosteroids and 2 received IVIg after plasma exchange. All but one of the CIDP patients were treated during follow-up, most received IVIg (18), the others received corticosteroids (6), plasma exchange (1), immunosuppressive drugs (2) or a combination of IVIg with corticosteroids and/or plasma exchange (6). Of the 9 MGUSP patients, 5 were treated: 2 with IVIg and 3 with rituximab. All MMN patients were treated with IVIg, and 1 had IVIg combined with cyclophosphamide.

Table 2. Characteristics of the longitudinal patients at entry and at 12 months

	GBS		CIDP		MGUSP		MMN	
	Baseline *n=36*	***12 months*** *n=11*	***Baseline*** *n=34*	***12 months*** *n=17*	***Baseline*** *n=9*	***12 months*** *n=3*	***Baseline*** *n=10*	***12 months*** *n=8*
Newly diagnosed: relapse; n:n	36 :0		19:15		6:3		5:5	
Gender; male:female, n:n	25:11	8:3	25:9	11:6	6:3	2:1	8:2	6:2
Age in years; median (range)	57 (19-90)	60 (39-91)	58 (18-74)	59 (32-75)	62 (45-85)	63 (47-86)	47 (32-60)	49 (35-61)
Grip strength (mean both hands)								
Jamar (pounds); median (range)	16 (0-99)	63 (21-135)	41 (0-116)	38 (5-99)	53 (31-109)	34 (5-61)	61 (14-103)	60 (19-75)
Vigorimeter (kPa); median (range)	32 (0-98)	74 (32-142)	55 (0-108)	61 (0-100)	74 (38-107)	59 (0-80)	77 (10-121)	68 (7-120)
ODSS arm grade; n (%)								
0	1 (3%)	9 (82%)	7 (21%)	6 (35%)			2 (20%)	1 (13%)
1	2 (6%)	1 (9%)	3 (9%)	2 (12%)	5 (56%)		1 (10%)	1 (13%)
2	11 (31%)	1 (9%)	12 (35%)	3 (18%)	4 (44%)	2 (67%)	4 (40%)	6 (75%)
3	9 (25%)		10 (29%)	5 (29%)			3 (30%)	
4	9 (25%)		2 (6%)			1 (33%)		
5	4 (11%)							
unknown				1 (6%)				
MRC sumscore; median (range)	48 (0-60)	60 (50-60)	52 (8-60)	58 (37-60)	57 (52-60)	58 (39-59)	131 (99-139)	131 (109-139)

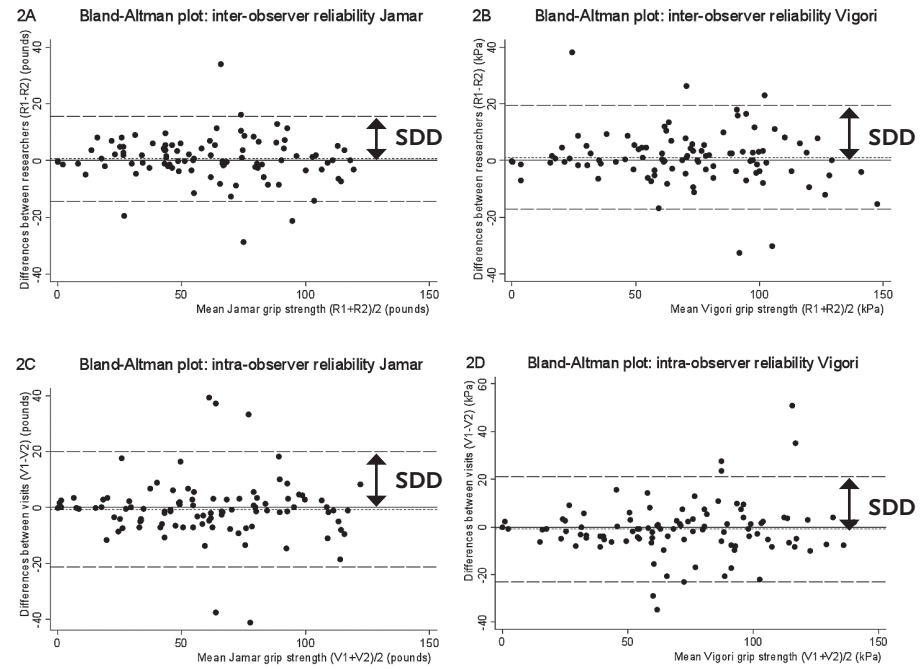
Floor and ceiling effects

Of the cross-sectional group 5 MMN patients reached a floor effect using the Jamar dynamometer and 3 when using the Vigorimeter. In the longitudinal group a floor effect was seen in 17 patients (13 GBS and 4 CIDP) using the Jamar dynamometer and in 18 patients (12 GBS and 6 CIDP) using the Vigorimeter. No ceiling effects were observed.

Validity and reliability

A strong correlation was found between the two grip strength devices (r_s= 0.86, p<0.0001). Correlation between both devices and the ODSS arm grade and the MRC sumscore were moderate and comparable (table 3). Similar intraclass correlation coefficients were found for both devices (table 3). Differences in grip strength values in a single investigator and between two investigators were approximately evenly distributed around the almost zero mean difference line of all Bland-Altman plots (figure 2). The SDD values derived from the intra-observer plots (20.5 pounds for the Jamar dynamometer and 22.0 kPa for the Vigorimeter) were higher than those derived from the inter-observer plots (15.0 pounds for the Jamar and 18.3 kPa for the Vigorimeter).

Figure 2: Bland-Altman plots reflecting inter-and intra-observer agreement for the Jamar (figure 2A+2C) and the Vigorimeter (figure 2B+2D). SDD= smallest detectable difference



Responsiveness

Correlations between the clinical judgment scores and the difference in grip strength values between visits and the calculated Guyatt responsiveness ratios for both devices are presented in table 3. During follow-up patients rated their clinical condition 64 times 'better compared to the last visit' and 39 times 'stable compared to the last visit' on the ClinJSc.

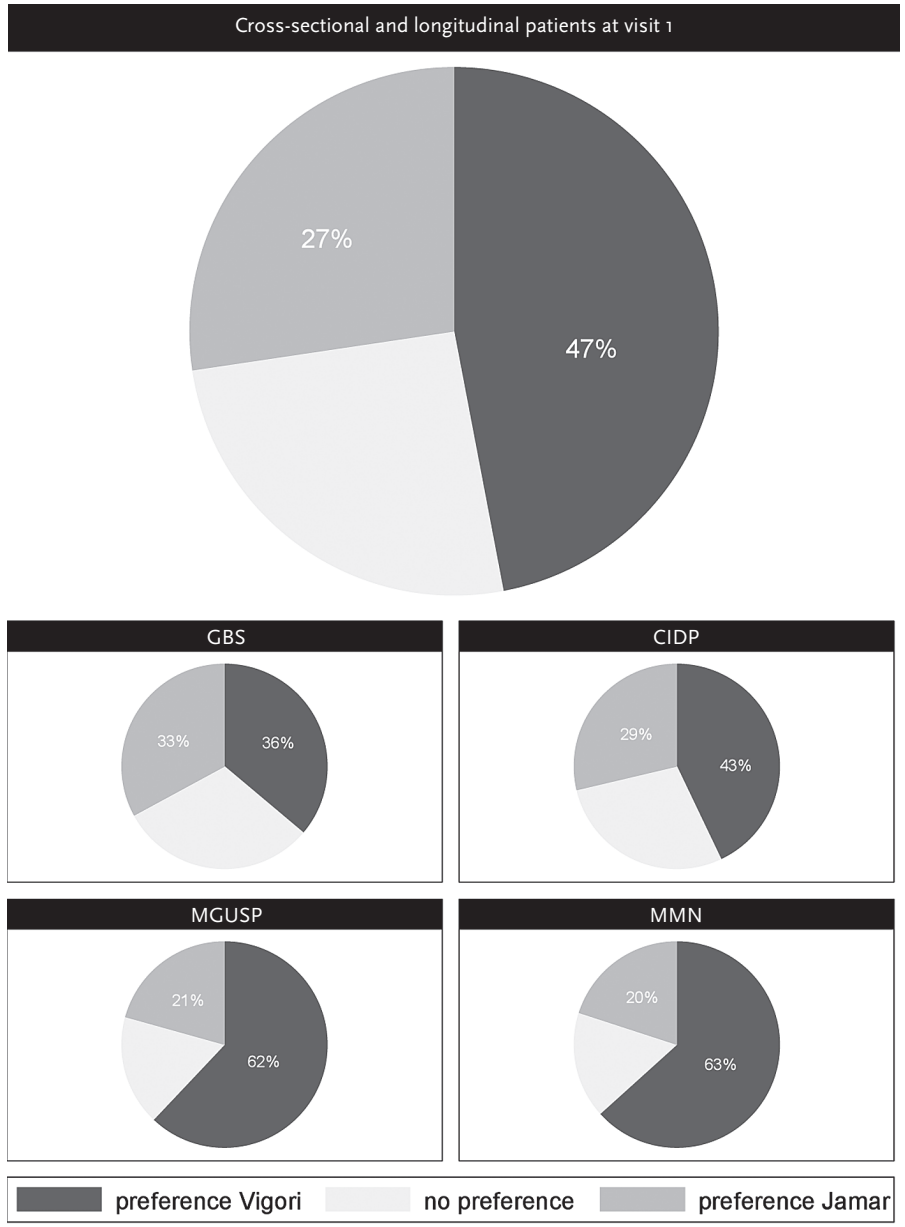
Table 3: Comparative validity, reliability and responsiveness of the Jamar and the Vigorimeter

	Jamar	Vigorimeter
Validity		
Spearman's rank correlation coefficient		
- correlation with the ODSS arm grades	-0.58	-0.55
- correlation with the MRC sumscores	0.59	0.58
Reliability		
Intraclass correlation coefficient (ICC) - intra-observer	0.996	0.946
Intraclass correlation coefficient (ICC) - inter-observer	0.956	0.977
Bland-Altman plot - systematic intra-observer difference	0.6 pound	1.0 kPa
Bland Altman plot - systematic inter-observer difference	0.7 pound	1.1 kPa
Responsiveness		
Spearman's rank correlation coefficient		
- correlation with Clinical judgment scores	0.23	0.43
Guyatt responsiveness ratio	0.64	0.56

Patients' preference

Cross-sectional patients consistently preferred the Vigorimeter above the Jamar dynamometer (Vigorimeter vs. Jamar; 49% vs. 26% at visit one, 55% vs. 29% at visit two). Also, 36 of the 59 longitudinal patients having a preference chose the Vigorimeter at the initial visit. The overall preference and the preference per condition are presented in figure 3. Patients' preference was not related to their grip strength values.

Figure 3: Patients' preference for grip strength device



Legend to figure 3. Per condition the preference of the cross-sectional and longitudinal patients summed.

Discussion

In conformity with earlier reports grip strength assessment with the Jamar dynamometer as well as the Vigorimeter demonstrated good validity and reliability.[4, 10-20] Our comparative validity analyses showed similar correlations between the obtained grip strength values of both devices with the ODSS arm grades and the MRC sumscores. Only one other study, in healthy children, directly compared reliability aspects of the Vigorimeter and a Jamar-like device (Lode dynamometer) and showing significantly lower intraclass correlation coefficients for the Vigorimeter.[34] In contrast, we observed high intraclass correlation coefficients and low systematic measurement errors (Bland-Altman plots) for both devices. These findings suggest that both devices are equally valid and reliable in immune-mediated neuropathies.

The ability to detect change in relation to an external anchor (clinical judgment score) was examined calculating Guyatt responsiveness ratios for both devices. A ratio greater than 1.96 is indicative of being a highly responsive device. However, both devices demonstrated similar low ratios suggesting limited responsiveness. In contrast, previous comparative responsiveness analyses in immune-mediated neuropathies ranked the Vigorimeter among the impairment and disability scales with the highest responsiveness capacity.[35] Another approach to select the device with the best chance of detecting true change is to consider the smallest detectable differences (SDDs) derived from the Bland-Altman plots. Differences greater than the SDD are necessary to detect a true change with 95% certainty for two consecutive grip strength assessments. Therefore, SDDs have been used to calculate the minimum clinically important difference (MCID), the smallest difference in score in the domain of interest which patients perceive as beneficial.[36] The calculated intra-observer SDDs (20.5 pounds for the Jamar dynamometer and 22 kPa for the Vigorimeter) were rather high compared to those in other populations.[10, 12, 34, 37] The difference in grip strength between baseline and 12 months was only in patients with GBS large enough to reach the SDDs of both devices. Further evaluation of MCID cut-off values will be necessary to ensure the clinical applicability of both grip strength devices in all immune-mediated neuropathies.

Our study demonstrated patients' preference for the Vigorimeter over the Jamar dynamometer. Reasons often mentioned for preferring the Vigorimeter include: 'less heavy', 'easier', 'more comfortable', 'less painful to squeeze a bulb' and 'squeezing makes you feel the force you are giving'. Those who preferred the Jamar dynamometer listed reasons like 'more robust', 'more suitable for big hands' and 'less chance of using it incorrectly'. Although normative values are available for both devices, the values calculated for the Jamar dynamometer do not consider the non-Gaussian distribution of grip strength values.[4, 17, 38] Therefore, we have collected Jamar grip strength values in healthy subjects to overcome this limitation

(see also chapter 2.2).[39] Based on our observations indicating equivalent validity, reliability and responsiveness for both devices, it seems that the Vigorimeter should be selected as the standard device for future studies based upon patients' preference.

Our study has some methodological limitations. Construct validity was calculated correlating linear grip strength scores with sumscores of ordinal outcome measures. Accordingly, the obtained ordinal sumscores were treated as if these were linear with all the components having equal relevance, which is highly unlikely.[40, 41] To calculate the Guyatt responsiveness ratio we used the clinical judgment score as an anchor. However, the correlation between this score and the difference in grip strength was only moderate for the Vigorimeter and even low for the Jamar dynamometer. If another anchor having higher correlations with difference in grip strength would be available, this could improve the responsiveness analyses. Furthermore, our sample size might be too small to draw a robust conclusion concerning responsiveness. Different disease courses of certain immune-mediated neuropathies may necessitate subgroup analyses to detect potential differences in responsiveness attributable to the underlying condition. The international multi-centre longitudinal part of the ***P*eri*N*om*S*** study will further evaluate these issues.

In conclusion, this study demonstrated that the Jamar dynamometer and the Vigorimeter show similar validity and reliablity to assess grip strength in immune-mediated neuropathies. Although responsiveness scores were rather low, this applied for both devices. Therefore, based on patients' preference, we recommend that the Vigorimeter should be used as the standard to assess grip strength for future studies in these conditions.

References

1. Bernsen, R.A., et al., *Residual health status after Guillain-Barré syndrome.* J Neurol Neurosurg Psychiatry, 1997. **62**(6): p. 637-40.
2. Rhind, V.M., H.A. Bird, and V. Wright, *A comparison of clinical assessments of disease activity in rheumatoid arthritis.* Ann Rheum Dis, 1980. **39**(2): p. 135-7.
3. Krischak, G.D., et al., *Physiotherapy after volar plating of wrist fractures is effective using a home exercise program.* Arch Phys Med Rehabil, 2009. **90**(4): p. 537-44.
4. Merkies, I.S., et al., *Assessing grip strength in healthy individuals and patients with immune-mediated polyneuropathies.* Muscle Nerve, 2000. **23**(9): p. 1393-401.
5. Sunderland, A., et al., *Arm function after stroke. An evaluation of grip strength as a measure of recovery and a prognostic indicator.* J Neurol Neurosurg Psychiatry, 1989. **52**(11): p. 1267-72.
6. Hahn, A.F., et al., *Intravenous immunoglobulin treatment in chronic inflammatory demyelinating polyneuropathy. A double-blind, placebo-controlled, cross-over study.* Brain, 1996. **119**(Pt 4): p. 1067-77.
7. Hughes, R., et al., *Randomized controlled trial of intravenous immunoglobulin versus oral prednisolone in chronic inflammatory demyelinating polyradiculoneuropathy.* Ann Neurol, 2001. **50**(2): p. 195-201.
8. Hughes, R.A., et al., *Intravenous immune globulin (10% caprylate-chromatography purified) for the treatment of chronic inflammatory demyelinating polyradiculoneuropathy (ICE study): a randomised placebo-controlled trial.* Lancet Neurol, 2008. **7**(2): p. 136-44.
9. Federico, P., et al., *Multifocal motor neuropathy improved by IVIg: randomized, double-blind, placebo-controlled study.* Neurology, 2000. **55**(9): p. 1256-62.
10. Espana-Romero, V., et al., *Elbow position affects handgrip strength in adolescents: validity and reliability of Jamar, DynEx, and TKK dynamometers.* J Strength Cond Res. **24**(1): p. 272-7.
11. Bellace, J.V., et al., *Validity of the Dexter Evaluation System's Jamar dynamometer attachment for assessment of hand grip strength in a normal population.* J Hand Ther, 2000. **13**(1): p. 46-51.
12. Smidt, N., et al., *Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis.* Arch Phys Med Rehabil, 2002. **83**(8): p. 1145-50.
13. Bohannon, R.W. and K.L. Schaubert, *Test-retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders.* J Hand Ther, 2005. **18**(4): p. 426-7, quiz 428.
14. Solgaard, S., B. Kristiansen, and J.S. Jensen, *Evaluation of instruments for measuring grip strength.* Acta Orthop Scand, 1984. **55**(5): p. 569-72.
15. Mathiowetz, V., *Comparison of Rolyan and Jamar dynamometers for measuring grip strength.* Occup Ther Int, 2002. **9**(3): p. 201-9.
16. Guerra, R.S. and T.F. Amaral, *Comparison of hand dynamometers in elderly people.* J Nutr Health Aging, 2009. **13**(10): p. 907-12.
17. Mathiowetz, V., et al., *Reliability and validity of grip and pinch strength evaluations.* J Hand Surg Am, 1984. **9**(2): p. 222-6.
18. Jones, E., et al., *Strength and function in the normal and rheumatoid hand.* J Rheumatol, 1991. **18**(9): p. 1313-8.
19. Lindstrom-Hazel, D., A. Kratt, and L. Bix, *Interrater reliability of students using hand and pinch dynamometers.* Am J Occup Ther, 2009. **63**(2): p. 193-7.
20. Rosen, B., L.B. Dahlin, and G. Lundborg, *Assessment of functional outcome after nerve repair in a longitudinal cohort.* Scand J Plast Reconstr Surg Hand Surg, 2000. **34**(1): p. 71-8.
21. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
22. van Nes, S.I., et al., *Peripheral Neuropathy Outcome Measures Standardisation (PeriNomS) study.* J Peripher Nerv Syst, 2008. **13**(2): p. 185.
23. *Research criteria for diagnosis of chronic inflammatory demyelinating polyneuropathy (CIDP). Report from an Ad Hoc Subcommittee of the American Academy of Neurology AIDS Task Force.* Neurology, 1991. **41**(5): p. 617-8.

24. Asbury, A.K. and D.R. Cornblath, *Assessment of current diagnostic criteria for Guillain-Barré syndrome.* Ann Neurol, 1990. **27 Suppl**: p. S21-4.
25. Miescher, G.C. and A.J. Steck, *Paraproteinaemic neuropathies.* Baillieres Clin Neurol, 1996. **5**(1): p. 219-32.
26. *European Federation of Neurological Societies/Peripheral Nerve Society Guideline on management of multifocal motor neuropathy. Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society.* J Peripher Nerv Syst, 2006. **11**(1): p. 1-8.
27. Casanova, J.S., *Clinical assessment recommendations.* 2nd ed. 1992, Chigago: The American Society of Hand Therapists.
28. Kleyweg, R.P., F.G. van der Meche, and P.I. Schmitz, *Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barré syndrome.* Muscle Nerve, 1991. **14**(11): p. 1103-9.
29. Van Asseldonk, J.T., et al., *Demyelination and axonal loss in multifocal motor neuropathy: distribution and relation to weakness.* Brain, 2003. **126**(Pt 1): p. 186-98.
30. Merkies, I.S., et al., *Clinimetric evaluation of a new overall disability scale in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2002. **72**(5): p. 596-601.
31. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.
32. Guyatt, G., S. Walter, and G. Norman, *Measuring change over time: assessing the usefulness of evaluative instruments.* J Chronic Dis, 1987. **40**(2): p. 171-8.
33. Bland, J.M. and D.G. Altman, *Multiple significance tests: the Bonferroni method.* Bmj, 1995. **310**(6973): p. 170.
34. Molenaar, H.M., et al., *Age-specific reliability of two grip-strength dynamometers when used by children.* J Bone Joint Surg Am, 2008. **90**(5): p. 1053-9.
35. Merkies, I.S., et al., *Comparison between impairment and disability scales in immune-mediated polyneuropathies.* Muscle Nerve, 2003. **28**(1): p. 93-100.
36. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status. Ascertaining the minimal clinically important difference.* Control Clin Trials, 1989. **10**(4): p. 407-15.
37. Nitschke, J.E., et al., *When is a change a genuine change? A clinically meaningful interpretation of grip strength measurements in healthy and disabled women.* J Hand Ther, 1999. **12**(1): p. 25-30.
38. Werle, S., et al., *Age- and gender-specific normative data of grip and pinch strength in a healthy adult Swiss population.* J Hand Surg Eur Vol, 2009. **34**(1): p. 76-84.
39. Peters, M.J.H., et al., *Revised normative values for grip strength with the Jamar dynamometer.* J Peripher Nerv Syst, 2011. **16**: p. 48-51.
40. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
41. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.

Chapter 4.2

Measuring sensory deficit in immune-mediated neuropathies:

comparing the sensory subset of the neuropathy impairment score with the modified INCAT sensory sumscore through Rasch analyses

Sonja I. van Nes[1], Els K. Vanhoutte[2], Mayienne Bakkers[2], Krista Kuitwaard[1], Kenneth C. Gorson[3], David R. Cornblath[4], Nicolette C. Notermans[5], Richard Lewis[6], Angelika Hahn[7], Guiseppe Lauria[8], Eduardo Nobile-Orazio[9], Grazia Devigili[10], Luca Padua[11], Jean-Marc Léger[12], Peter Y.K. van den Bergh[13], David Bennett[14], Robert Hadden[14], Amilton A. Barreira[15], Pieter A. van Doorn[1], Catharina G. Faber[2], Ingemar S.J. Merkies[1,2,16], on behalf of the PeriNomS study group

[1]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of Neurology, Maastricht University Medical Centre, Maastricht, the Netherlands
[3]Department of Neurology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, USA
[4]Department of Neurology, Johns Hopkins Hospital and School of Medicine, Baltimore, USA
[5]Department of Neurology, Rudolf Magnus Institute of Neuroscience University Medical Centre, Utrecht, the Netherlands
[6]Department of Neurology, Wayne State University School of Medicine, Detroit, USA
[7]Department of Neurology, London Health Science Center, London, Ontario, Canada
[8]Department of Neurology, Neuromuscular Diseases Unit, National Neurological Institute Carlo Besta, Milan, Italy
[9]Department of Neurological Sciences, Milan University, Clinico Humanitas, Milan, Italy
[10]Neurological Unit Ospedale dell-Angelo, Mestre Venezia, Italy
[11]Department of Neurology, Università Cattolica, Rome, Italy
[12]Department of Neurology, Hôpital Pitié Salpêtrière, Paris, France
[13]Department of Neurology, Catholique University of Louvain, Belgium
[14]Department of Neurology, King's college Hospital NHS Foundation Trust, London, United Kingdom
[15]Department of Neurosciences, Hospital das Clínicas de Ribeirão Preto, Ribeirão Preto, Brazil
[16]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands

Manuscript in preparation

Abstract

Background: Although the sensory subset of the neuropathy impairment score (NISs) and the INCAT sensory sumscore (ISS) both fulfil all clinimetric requirements in patients with immune-mediated neuropathies, it is unknown which scale is superior.

Objective: Comparison of the NISs and the modified ISS (mISS) to select one sensory scale as the standard for future follow-up studies and clinical trials in patients with Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), and monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP).

Methods:
Step 1. NISs and mISS data were subjected to Rasch analyses to transform these ordinal sensory scales to interval scales (1st visit data of 82 newly diagnosed patients plus 1st and 2nd visit data of 80 clinically stable patients).
Step 2. Comparing clinimetric properties of the (Rasch-built) NISs and mISS. Construct validity and explanatory validity (correlation with disability) were calculated for both scales. To determine reliability both scales were assessed twice (interval 2-4 weeks) in the clinically stable patients (30 GBS, 30 CIDP, 20 MGUSP). Finally, traditional group-level responsiveness (effect size (ES)) and individual-person-level responsiveness (MCID-SE: individual change/standard error of difference (SE_{diff})) were estimated for each scale in 74 newly diagnosed patients (32 GBS; 33 CIDP, 9 MGUSP).

Results: A Rasch-built interval scale could be constructed for both sensory measures demonstrating good fit statistics, with the Rasch-built mISS (R-mISS) demonstrating a broader span than the Rasch-built NISs (R-NISs); R-mISS: 5.01 logits vs. R-NISs: 3.85 logits. Construct validity (R^2=0.72; p<0.0001) was good. However, explanatory validity was poor (mISS: 22% and NISs 18% of disability explained). Reliability aspects were good and comparable. Responsiveness values (ES and MCID-SE) were constantly higher for the (R-)mISS compared to the (R-) NISs.

Conclusion: The mISS and NISs were successfully transformed to interval scales with a linear construct using the Rasch model. Both scales demonstrated similar satisfactory clinimetric findings. However, the R-mISS had a larger targeting range, less floor effect and seemed to have slightly better responsiveness scores than the R-NISs. Therefore, the use of the R-mISS is suggested for future studies in immune-mediated neuropathies.

Introduction

Sensory deficit may contribute to limitations in activity and participation in patients with immune-mediated neuropathies. The sensory subset of the neuropathy impairment score (NISs) and the Inflammatory Neuropathy Cause and Treatment (INCAT) sensory sumscore (ISS) have been widely used to assess sensory deficit in these conditions. Both scales have demonstrated good traditional scientific soundness, including responsiveness.[1-4] However, it is unknown which scale is superior.

At a workshop on outcome measures, both scales were criticised due to their apparent deficiencies.[5] The NISs embraces touch pressure, pinprick, joint position and vibration sense, but sensory deficit is being assessed only distally at the index finger and great toe, hereby omitting proximal sensory disturbances.[2, 3, 5] In addition, vibration sense is being assessed using a non-graduated 165 Hz tuning fork, hence not taking into account the possible graduation of this quality at bedside.[6] Furthermore, assessment of two-point discrimination at the index finger is not incorporated in the NISs.

The original ISS measures pinprick and vibration sense in the arms and legs from distal to proximal recording the highest extension of dysfunction, plus two-point discrimination at the index finger.[1] Joint position and light touch modalities however are not incorporated, making this scale less comprehensive and perhaps less responsive.[5] Additionally the two-point discrimination categories were arbitrarily taken.[7] Therefore, we created the modified ISS (mISS) combining the apparent advantages of both scales. The modalities touch pressure and joint position were incorporated and proximal assessment was preserved. In addition, static two-point discrimination values were collected and graded as normal or abnormal based on age-related reference values now available (see also chapter 2.1).[8] Still, the biggest disadvantage of both the NISs and the (m)ISS is that both scales are ordinal composite measures. The obtained scores are treated as if these were linear with all modalities having equal relevance, which is highly unlikely.[9, 10] Therefore, we investigated whether the NISs and the mISS could be translated to interval scales showing the exact distance between patients' ability levels and items' difficulty levels on a linear ruler using the Rasch method.[11, 12] In addition, from the constructed interval measures, we compared clinimetric qualities with emphasis on responsiveness in order to determine whether one scale is superior to the other.[4, 13]

Methods

Patients

Cross-sectional group (for Rasch modelling, validity, and reliability studies)

Eighty patients (30 patients with Guillain-Barré syndrome (GBS), 30 with chronic

inflammatory demyelinating polyradiculoneuropathy (CIDP), and 20 patients with monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP), aged 18 years and older with a stable clinical condition were recruited for the purposes of the current study (recruitment period: July 2008 to November 2010). A stable clinical condition was defined as an unchanged clinical functionality as declared by the patient to the best of his knowledge, plus (if applicable) no objective changes at neurological examination when compared with previous clinical findings over two months before the start of the study. All patients were recruited from three university outpatient clinics in the Netherlands (Erasmus University Medical Centre of Rotterdam, University Medical Centre Utrecht and Maastricht University Medical Centre).

Longitudinal group (for Rasch modelling and responsiveness studies)

Eighty-two patients with newly diagnosed GBS (36), CIDP (35), or MGUSP (11) were enrolled. These patients are participants in the on-going international multi-centre longitudinal part of the *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study in immune-mediated neuropathies (37 from the Netherlands, 18 from the USA, 11 from Italy, 5 from Belgium, 4 from Canada, 2 from France, 2 from the United Kingdom, 1 from Spain, and 2 from Brazil). All patients with GBS, all but one of the patients with CIDP and five of the 11 patients with MGUSP were treated according to standard regimens (e.g., with intravenous immunoglobulin, plasma exchange, corticosteroids, rituximab)

All patients with GBS and CIDP met the international criteria for their illness.[14, 15] All patients with MGUSP had IgM anti-MAG antibodies, with demyelinating features in nerve conduction studies. The diagnosis of MGUSP was established after excluding all other possible causes for gammopathy and polyneuropathy.[16]

Outcome measures

- *Neuropathy impairment score – sensory subset (NISs):* The NISs was applied according to the published procedures and comprises touch pressure, pinprick, vibration sense and joint position.[2, 3] In brief, touch pressure was assessed with long fibre cotton wool, pinprick with straight pins and vibration sense with a non-graduated 165 Hz tuning fork on the dorsal surface of the terminal phalanx of the index finger and great toe on both sides. Joint position was assessed by moving the terminal phalanx of the index finger and great toe on both sides.[2, 3] Each modality was scored as being normal (0), decreased (1), or absent (2). The NISs ranges from 0 (no sensory deficit) to 32 (most severe sensory deficit).
- *Modified INCAT sensory sumscore (mISS):* Based on recommendations given at a workshop on outcome measures in inflammatory neuropathies, we modified the original ISS.[5] The mISS embraces light touch, pinprick, vibration sense, joint position and two-point discrimination assessment. Light touch and pinprick modalities were assessed using disposable cotton wools and pins. Vibration sense was assessed using a graduated 64 Hz Rydel Seiffer tuning fork and graded based on published normative values.[6] Joint position was assessed in a standardised manner according to previous recommendations.[17, 18] In addition, a static two-point discrimination value was collected using a sliding aesthesiometer. Two-point discrimination was assessed only at the right index finger, all other modalities were assessed from distal to proximal in the arms and legs, with only the most affected side being recorded.1 Modalities were graded as normal (grade 0) or disturbed; at the index finger or hallux (grade 1); at the wrist or ankle (grade 2), at the elbow or knee (grade 3), at shoulder or hip joint (grade 4). Two-point discrimination was graded as normal (grade 0) or disturbed (grade 1) based on age-dependent normative values.[8] The mISS ranges from 0 (no sensory deficit) to 33 (most severe sensory deficit).
- *Rasch-built overall disability scale (R-ODS):* This scale is a 24-item interval measure specifically designed to capture activity limitations and participation restrictions in patients with GBS, CIDP, and MGUSP.[19] The R-ODS summed raw scores range from 0 (most severe activity and social participation restriction) to 48 (no activity and social participation restrictions). Without missing data these raw scores can be translated to logits demonstrating the exact distance between patients' ability levels on a linear scale, ranging from -6.95 (most severe activity and social participation restriction) to 8.11 (no activity and social participation restrictions) logits.

Study design

Medical ethics, patient consent and training

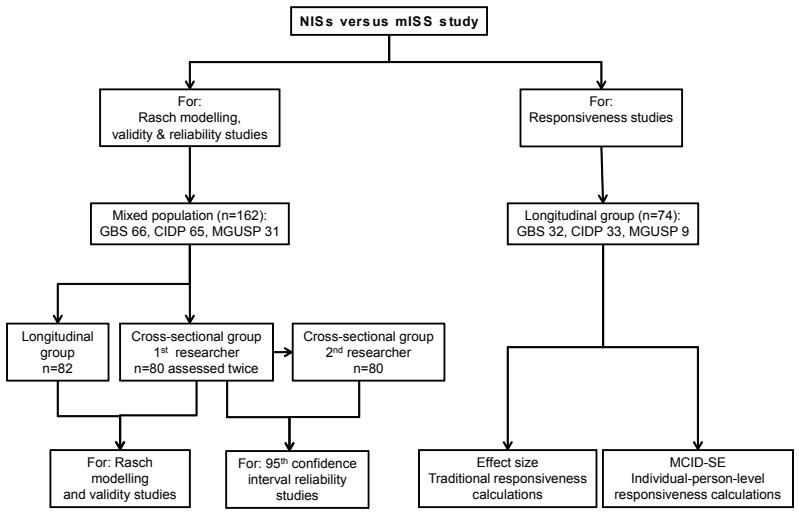
The local medical ethics committee of all participating centres approved the protocol. Written informed consent was obtained from each participant. All investigators participating in the international ***P*eri*N*om*S*** study (n=26) were trained in 2007 (by IM and SvN) at the Peripheral Nerve Society meeting (Utah, USA) aiming to standardise the assessment procedures for all scales as part of this study. In addition, participants received a comprehensive research manual that included a thorough description and pictures illustrating how to assess the various sensory modalities for the mISS. For the NISs, a video showing in practice how to measure the various entities was provided.

Examination schedule

All 80 cross-sectional patients were examined twice by the same investigator with an interval of 2-4 weeks, both times assessing the NISs, the mISS and the R-ODS. During one of those two visits a second investigator also assessed the NISs and the mISS for inter-rater reliability studies. For responsiveness analyses follow-up

records of 74 longitudinally studied patients (32 GBS, 33 CIDP and 9 MGUSP) were available. Patients with GBS and CIDP were investigated at entry, 1, 3, 6, and 12 months of follow-up. Patients with MGUSP were examined at entry, 3, and 12 months (see figure 1 for the study algorithm).

Figure 1. Study algorithm



Legend to figure 1. NISs = neuropathy impairment score - sensory subset. mISS = modified INCAT sensory sumscore. First visit data of the cross-sectional and the longitudinal group (n=162) and the assessment findings of the second visit of the cross-sectional sample (n=80) were stacked (162 + 80 = 242) to strengthen the Rasch model.

Statistical analyses - step 1: From ordinal to interval measures using Rasch

The NISs and mISS were subjected to Rasch analyses to determine whether model expectations would be met. The Rasch method including its expectations (e.g., proper fit statistics, unidimensionality, no item bias or local dependency) has been described thoroughly elsewhere, also specifically for neurologists (see also chapter 3.1).[11, 12, 19] In brief, this statistical technique attempts to transform ordinal obtained scores that are scale dependent and of limited accuracy, into interval measures that are scale independent and suitably accurate for individual patient assessment. In essence, this method is based on a logical assumption: individuals that are clinically less affected should have an increased probability, relative to clinically more affected individuals, of getting a better score.[11, 12, 19] First visit NISs and mISS data of the cross-sectional and the longitudinal group (n=162) and second visit data of the cross-sectional sample (n=80) were stacked (162 + 80 = 242) to strengthen the Rasch model, hereby controlling for 'time factor' as possible confounder and

fulfilling the minimum requirement for scale stability.[20, 21] Personal factors like age, gender, illness, and country were collected. For the purposes of the current study, these factors were categorised as follows: age (<50 years vs. 50-59 years vs. 60-69 years vs. ≥70 years), gender (female vs. male), illness (GBS vs. CIDP vs. MGUSP), and country (the Netherlands vs. others). Age was categorised as such striving for an equivalent distribution of participants per subgroup. Analyses were performed to obtain an interval scale for both sensory measures separately. Attempts were made to maintain the structures of the NISs and mISS while meeting the Rasch model expectations.

Statistical analyses - step 2: Clinimetric comparison of the NISs and the mISS

Validity and reliability studies

Construct convergent validity was obtained by correlating the scores of the obtained interval measures, the Rasch-built NISs (R-NISs) and the Rasch-built mISS (R-mISS), using quantile regression with restricted cubic spline function to overcome possible non-Gaussian distribution of the data.[22, 23] Explanatory validity (sensory deficit explaining disability) was determined by correlating both the R-NISs and the R-mISS scores with the first visit R-ODS scores in both patient groups. Internal reliability was examined by determining the Person Separation Index (PSI) for each of the interval sensory scales, separately. In general, a PSI above 0.7 is considered acceptable, indicating the ability to identify at least two groups of patients.[24] In addition, reliability studies were performed comparing the first and second assessment data of the cross-sectional sample to investigate whether hierarchy of patients' ability locations were consistent over time.[25] Validity and reliability were quantified by calculation of the intra-class correlation coefficient using a one-way random effects analysis of variance (ANOVA) model.

Responsiveness studies

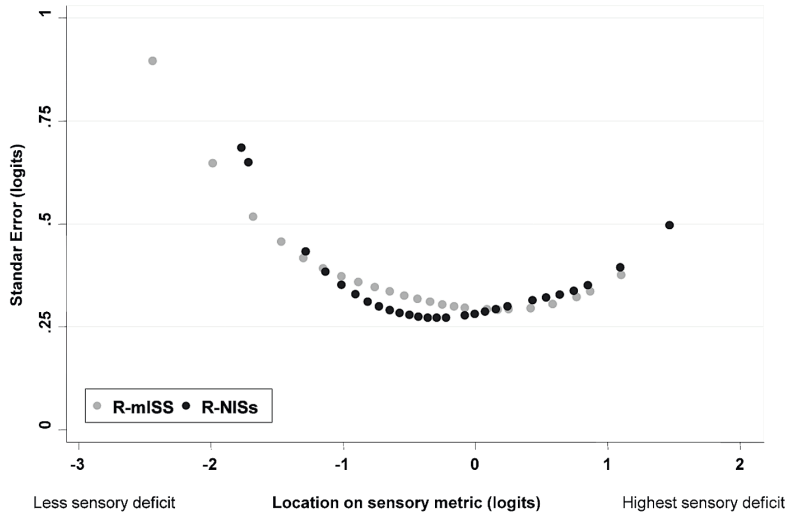
Traditional responsiveness at *group level* was calculated using the effect size (ES) indicator at 1, 3, 6 and 12 months of follow-up using the traditional sumscore values for both sensory scales. The ES is equal to the mean change in score divided by the standard deviation of the scores at entry ($ES=\mu i-\mu o/SD_{entry}$; μi = mean NISs or mISS sumscore of the longitudinally studied group at month = *i*; μo = mean NISs or mISS sumscore at entry).[26] According to Cohen's rule of thumb, an effect size value between 0.5 and 0.8 is considered moderate, and ≥ 0.8 represents high responsiveness.[27] The Rasch method demonstrates that the error around an individual personal score (standard error (SE)) may vary across the theoretical range of an outcome measure, and with this also the significance of individual changes (see also figure 2).[28] Therefore, responsiveness at the *individual-person-level* was calculated and referred to as minimum clinically important difference-standard error (MCID-SE) score based on the previously described significant change

(SigChange) score.[29] In brief, MCID-SE scores were calculated by computing for each of the 74 longitudinally studied patients separately: a) their individual change (person location at *i* month minus person location at entry; where *i* = 1, 3, 6, or 12 months of follow-up for patients with GBS and CIDP and *i* = 3 or 12 months for patients with MGUSP), and, b) the corresponding SE of difference related to their locations at entry and at *i* months (SE_{diff} = square-root (SE_{entry}^2+$SE_{i\ month}^2$), and c) the final MCID-SE calculations by dividing the individual change scores by corresponding SE_{diff} (MCID-SE = individual change/SE_{diff}).[29] Since the changes in sensory deficits tend to be small in these disorders, we have arbitrarily defined the minimum clinically important difference (MCID) cut-off using a change of at least 1 SE (69%CI).[30-32] Accordingly, the obtained MCID-SE scores were classified into the following subgroups:

- subgroup 1 (clinically important deterioration): MCID-SE ≥ 1;
- subgroup 2 (clinically unimportant deterioration): 0 < MCID-SE < 1;
- subgroup 3 (no change): MCID-SE = 0;
- subgroup 4 (clinically unimportant improvement): -1 < MCID-SE < 0;
- subgroup 5 (clinically important improvement): MCID-SE ≤ -1

The distribution of MCID-SE subgroups among the longitudinally studied patients for both sensory measures was compared using chi square statistics.

Figure 2. Graph showing the changing standard error with changing patients' ability locations on the sensory metrics



Legend to figure 2. A 'U'-shape pattern is shown indicating that the standard error changes with the changing patients' ability locations on the theoretical range of the R- mISS or the R-NISs. R- mISS = Rasch-built modified INCAT sensory sumscore, R-NISs = Rasch-built sensory subset of the neuropathy impairment score.

Statistics and software

Rasch analyses were performed with the partial credit model as default (RUMM2030). Further statistical analyses were undertaken using Stata 11.0 for Windows XP. The p-value was adjusted throughout the analyses, based on Bonferroni multiple testing corrections.[33]

Results

General aspects

The general characteristics of participants are presented in table 1.

Table 1. General characteristics of participants

	Cross-sectional group (for Rasch modelling, validity and reliability studies)	Longitudinal group (for Rasch modelling and responsiveness studies)
Number of patients	80	82
Age (years); mean (SD), range	62 (10), 26-81	55 (16), 18-90
Gender; n (%) female male	 29 (36) 51 (64)	 25 (31) 57 (70)
Diagnosis; n (%) GBS CIDP MGUSP	 30 (38) 30 (38) 20 (25)	 36 (44) 35 (43) 11(13)
R-ODS (logits); mean (SD), range	3.04 (2.13), -2.79-6.55	-.61 (3.47), -6.95-5.71
Country of assessment; n (%) the Netherlands USA Italy Belgium Canada France UK Spain	 80 (100)	 34 (46) 17 (23) 10 (14) 5 (7) 4 (5) 2 (3) 1 (1) 1 (1)

Step 1: From ordinal to interval measures using Rasch

Initial Rasch analyses on the sensory measures

Misfit to the Rasch model was demonstrated for both sensory measures (table 2, Initial analysis). The person fit residuals showed acceptable fit statistics, whereas the obtained scores (particularly the SD) for the item fit residuals deviated from the expected value of 1 (mISS: SD 1.697, NISs: 1.501). The significant chi square probability for both measures demonstrated no invariance of items. A substantial proportion of the t-tests performed fell outside the ± 1.96 range, indicating multidimensionality (table 2).

Data handling to fit the Rasch model

Throughout the analyses, we continuously monitored the distribution of persons within the class intervals, the overall fit statistics and the independent tests for unidimensionality. The aim was to modify the data of each scale to fulfil model requirements maintaining the structure of the original scale. The 'time factor' introduced because of stacking the data was not a confounding factor for both sensory measures.

- *NISs:* All eight sensory modality pairs (e.g., vibration sense on the left and the right great toe, joint position on the left and right index finger) demonstrated a strong local dependency (ρ>0.696-0.924). Subsequently, eight subsets of items were created combining left and right scores of each modality. Touch pressure and pinprick still demonstrated local dependency for the fingers as well as the toes. Based on these findings and the anatomical distribution of the sensory modalities, a second round of subsets was created combining touch pressure with pinprick and vibration sense with joint position creating 4 subsets (vital sense finger, proprioceptive sense finger, vital sense toe, proprioceptive sense toe). After this, model expectations were met, reflected by good fit statistics, no local dependency and unidimensionality (table 2, final analysis). There was one uniform item bias (on personal factor age category for vital sense in the legs), which was accepted in order to maintain the internal structure of the NISs. A total of 9.9% of the patients (24/242) demonstrated a floor effect and 0.8% had a ceiling effect (2/242). The R-NISs ranged from -1.76 to 2.09 (3.85) logits.

- *mISS:* Similar to the NISs, correlations were seen between the vital modalities in the arms and legs, separately. Therefore, similar subsets were created also incorporating two-point discrimination into the subset proprioceptive sense arms. After this, the R-mISS fulfilled the Rasch model requirements and was free of item bias and local dependency, and demonstrated good unidimensionality (table 2, final analyses). A total of 8.7% (21/242) of the patients demonstrated a floor effect. No ceiling effect was seen. The R-mISS ranged from -2.43 to 2.58 (5.01) logits.

Step 2 - Clinimetric comparison of the NISs and mISS

Validity and reliability studies

The R-NISs and R-mISS demonstrated good correlation with each other (construct validity: R^2=0.72; p<0.0001). However, explanatory validity (sensory deficits explaining disability as measured with the R-ODS through variance studies) was poor: the NISs explained 18% of the disability findings; for the mISS this was slightly higher (22%). Good person separation indices (PSI) were obtained for both sensory measures (table 2, final analyses). The test-retest reliability of items' difficulty locations was good (NISs: R^2=0.9; mISS: R^2=0.94). Also, patients' ability locations were almost always within the 95% confidence interval lines, reflecting acceptable test-retest reliability (figure 3).

Table 2. Summary statistics of Rasch analyses, for (A) the mISS and (B) the NISs

A. modified INCAT sensory sumscore

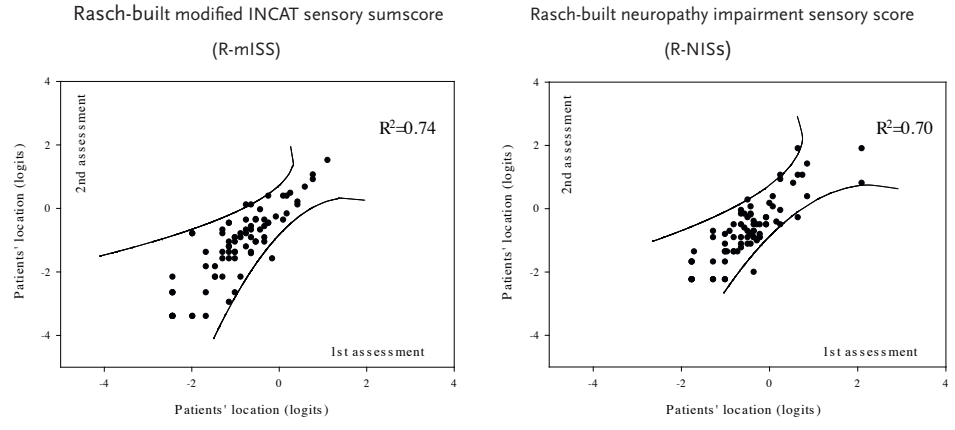
Rasch analyses	Item Fit Residuals		Person Fit Residuals		Item-trait Chi-square interaction		PSI	Unidimensionality indep. t-test (95%-CI)
	Mean	SD	Mean	SD	DF	p-value		
Initial	0.209	1.697	-0.318	1.037	27	0.00001	0.87	0.10 (0.07-0.12)
Final	0.209	0.671	-0.393	0.940	12	0.383	0.80	0.03 (0.003-0.06)

B. Neuropathy impairment score – sensory subset

Rasch analyses	Item Fit Residuals		Person Fit Residuals		Item-trait Chi-square interaction		PSI	Unidimensionality Indep. t-test (95%-CI)
	Mean	SD	Mean	SD	DF	p-value		
Initial	-0.456	1.501	-0.361	0.949	48	0.000006	0.94	0.20 (0.18-0.23)
Final	0.113	1.175	-0.237	0.795	12	0.053	0.84	0.03 (0.003-0.06)

Legend to table 2. CI = confidence interval, DF = degrees of freedom, PSI = person separation index, SD = standard deviation.

Figure 3. Test-retest reliability of patients' ability locations for the Rasch-built NISs and Rasch-built mISS

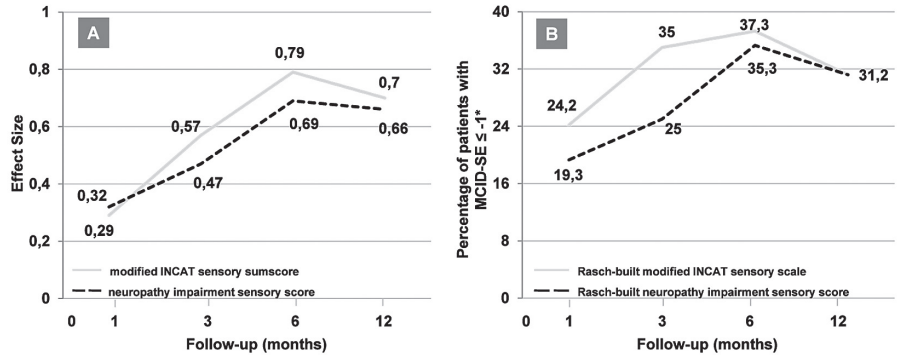


Legend to figure 3. Patients' ability locations in the first visit versus the second visit. Almost all persons (dots) were located within the 95% confidence interval (solid lines).

Responsiveness studies

Traditional responsiveness scores using the effect size (ES) varied from poor to moderate for both scales. Slightly better responsiveness scores were seen for the mISS (figure 4A). The responsiveness calculation at the *individual-person-level*, taking the changing standard errors into account (see also figure 2), demonstrated higher percentages of patients with clinically important improvement on the Rasch-built mISS, especially during the first months of follow-up (figure 4B). Compared to the Rasch-built NISs also more patients with clinically important deterioration were seen on the Rasch-transformed mISS during the first month (table 3). The differences were however not significant.

Figure 4. Responsiveness findings (figure 4A. effect size and figure 4B. MCID-SE) in longitudinally studied patients with GBS, CIDP and MGUSP



Legend to figure 4. (B) *patients with MCID-SE<=1 correspond to subgroup 5 (clinically important improvement)

Table 3. Percentages of longitudinally studied patients per subgroup for the R-NISs and the R-mISS, categorised based on their MCID-SE cut-off values during follow-up

Subgroup classification	Rasch-built NISs				Rasch-built mISS			
	% of longitudinal studied patients during follow-up (months)				% of longitudinal studied patients during follow-up (months)			
	1	3	6	12	1	3	6	12
1 = clinically important deterioration (MCID-SE ≥1)	8.1%	6.7%	3.9%	6.3%	8.1%	1.6%	2.0%	3.1%
2 = clinically unimportant deterioration (0 < MCID-SE < 1)	14.5%	8.3%	11.8%	3.1%	14.5%	21.7%	7.8%	0.0
3 = no change (MCID-SE = 0)	21.0%	11.7%	9.8%	12.5%	16.1%	10.0%	13.7%	15.7
4 = clinically unimportant improvement (-1 < MCID-SE < 0)	37.1%	48.3%	39.2%	46.9%	37.1%	31.7%	39.2%	50.0
5 = clinically important improvement (MCID-SE ≤ -1)	19.3%	25.0%	35.3%	31.2%	24.2%	35.0%	37.3%	31.2

Legend to table 3. MCID-SE = minimum clinically important difference-standard error. The MCID-SE refers to individual-person level responsiveness calculated by individual change divided by their corresponding SE of difference (related to their locations at entry and at *i* months where *i* = 1, 3, 6, or 12 months of follow-up for patients with GBS and CIDP and *i* = 3 or 12 months for patients with MGUSP).
Note: A negative MCID-SE score corresponds to an improvement, since the original sensory scales range from 0 (no deficit) to 32 and 33 (maximum sensory deficit), for the NISs and mISS respectively.

Discussion

This study compared two widely used composite sensory scales in patients with immune-mediated neuropathies, the NISs and the mISS.[1-3] First, both scales were subjected to Rasch analyses which demonstrated that the whole may not equal the sum of the parts for the NISs and mISS as ordinal composite measures.[10] Both scales did not meet the Rasch model expectations showing in particular local dependency between the various sensory modalities. After creating subsets acceptable model fulfilment was obtained for the mISS and the NISs, hereby constructing two sensory interval measures.

Secondly, both scales were compared regarding their clinimetric properties. At first glance, the two constructed sensory interval measures demonstrated similar clinimetric properties, despite the differences regarding location of assessment (NISs measuring only at the index finger and great toe; in contrast, mISS capturing sensory deficit up to the shoulders and hips). However, targeting of the R-mISS was better, with a larger span (5.01 logits) when compared to the span of the R-NISs (3.85 logits). This probably explains the lower floor effect and no ceiling effect for the R-mISS. In addition, the responsiveness calculations, both effect size and MCID-SE, were in favour of the (R-)mISS. Its responsiveness scores were generally higher throughout the follow-up period and captured clinically important improvement and clinically important deterioration faster (figure 4 and table 3). Apparently, the (R-)mISS had a higher ability to capture clinically important changes in patients with GBS, CIDP, and MGUSP. Based on these observations, a preference may be given to the use of the Rasch-built mISS.

The Rasch method demonstrated the varying standard errors across the theoretical range of the interval sensory measures, and with this also the significance and direction of clinically important changes in patients with different ability levels (Table 3).[28, 29] These findings are in contrast with traditional methods, stating that the measurement error around a patient is a constant value.[29] The Rasch method enabled the use of individual standard errors based on patient ability levels. However, we arbitrarily defined a change of 1xSE a minimum clinically important difference (MCID), thereby accepting that some change due to measurement error might be attributed to clinically important change. A higher cut-off value, like 1.96xSE, reduces the percentage of patients categorised as being a responder, but more importantly defines small changes as clinically unimportant which is arguable.

In conclusion, we succeeded in constructing interval measures for the NISs and the mISS in patients with GBS, CIDP and MGUSP by subjecting both ordinal sensory scales to Rasch analyses. Despite the apparent similarity, the clinimetric comparative studies were in favour of the R-mISS. This scale demonstrated a larger targeting range, less floor effect, and seems to have better responsiveness scores. Therefore, the use of the Rasch-built mISS is suggested, as part of the final aim of the ***P*eri*N*om*S*** study, to standardise outcome measures applied in future clinical trials and observational studies in patients with GBS, CIDP, and MGUSP.

References

1. Merkies, I.S., et al., *Psychometric evaluation of a new sensory scale in immune-mediated polyneuropathies. Inflammatory Neuropathy Cause and Treatment (INCAT) Group.* Neurology, 2000. **54**(4): p. 943-9.
2. Dyck, P.J., et al., *The Rochester Diabetic Neuropathy Study: design, criteria for types of neuropathy, selection bias, and reproducibility of neuropathic tests.* Neurology, 1991. **41**(6): p. 799-807.
3. Dyck, P.J., et al., *The Rochester Diabetic Neuropathy Study: reassessment of tests and criteria for diagnosis and staged severity.* Neurology, 1992. **42**(6): p. 1164-70.
4. Streiner, D.L. and G.R. Norman, *Health measurement scales. A practical guide to their development and use* 2nd ed. 1998, New York: Oxford University Press.
5. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
6. Martina, I.S., et al., *Measuring vibration threshold with a graduated tuning fork in normal aging and in patients with polyneuropathy. European Inflammatory Neuropathy Cause and Treatment (INCAT) group.* J Neurol Neurosurg Psychiatry, 1998. **65**(5): p. 743-7.
7. Dyck, P.J., R.A.C. Hughes, and P.C. O'Brien, *Quantitating overall neuropathy symptomes, and outcomes*, in *Peripheral Neuropathy*, P.J. Dyck and P.K. Thomas, Editors. 2005, WB Saunders Company: Philadelphia.
8. van Nes, S.I., et al., *Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies.* J Neurol Neurosurg Psychiatry, 2008. **79**(7): p. 832-4.
9. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
10. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.
11. Rasch, G., *Probabilistic models for some intelligence and attainment tests.* 1960, Copenhagen: Danmarks Paedagogiske Institut.
12. Tennant, A. and P.G. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* Arthritis Rheum, 2007. **57**(8): p. 1358-62.
13. Feinstein, A.R., *Clinimetrics*. 1987, New Haven and London: Yale University Press.
14. *Research criteria for diagnosis of chronic inflammatory demyelinating polyneuropathy (CIDP). Report from an Ad Hoc Subcommittee of the American Academy of Neurology AIDS Task Force.* Neurology, 1991. **41**(5): p. 617-8.
15. Asbury, A.K. and D.R. Cornblath, *Assessment of current diagnostic criteria for Guillain-Barre syndrome.* Ann Neurol, 1990. **27 Suppl**: p. S21-4.
16. Miescher, G.C. and A.J. Steck, *Paraproteinaemic neuropathies.* Baillieres Clin Neurol, 1996. **5**(1): p. 219-32.
17. Gilman, S., *Joint position sense and vibration sense: anatomical organisation and assessment.* J Neurol Neurosurg Psychiatry, 2002. **73**(5): p. 473-7.
18. Stillman, B., *Thesis: an investgation of the clinical assessment of joint position sense*, in *School of physiotherapy*. 2000: Victoria, Australia.
19. van Nes, S.I., et al., *Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies.* Neurology, 2011. **76**(4): p. 337-45.
20. Linacre, J., *Sample size and item calibration stability.* Rasch Meas Trans, 1994. **7**: p. 328.
21. Wright, B.D., *Rack and stack: Time 1 vs Time 2.* Rasch Meas Trans, 2003. **17**: p. 905-906.
22. Gould, W.W. and W.H. Rogers, *Quantile regression as an alternative to robust regression. Proceedings of the statistical computing section.* 1994, Alexandria; Virginia: American Statistical Association
23. Herndon J.E., H.F.E.J., *The restricted cubic spline hazard model.* Comm Stat Theory Meth, 1990. **19**: p. 639-663.
24. Fischer, W.P., *Reliability statistics.* Rasch Meas Trans, 1992. **6**: p. 238.
25. Wright, B.D. and M.H. Stone, *Best test design*. 1979, Chicago: Mesa Press.
26. Kazis, L.E., J.J. Anderson, and R.F. Meenan, *Effect sizes for interpreting changes in health status.* Med Care, 1989. **27**(3 Suppl): p. S178-89.

27. Cohen, J., *Statistical power analysis for the behavioural sciences* L. Erlbaum, Editor. 1988: New York. p. p 24–25.
28. Lai, J.S. and D.T. Eton, *Rasch Meas Transact.* 2002. **15**: p. 850.
29. Hobart, J. and S. Cano, *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.* Health Technol Assess, 2009. **13**(12): p. iii, ix-x, 1-177.
30. Hughes, R.A., et al., *Intravenous immune globulin (10% caprylate-chromatography purified) for the treatment of chronic inflammatory demyelinating polyradiculoneuropathy (ICE study): a randomised placebo-controlled trial.* Lancet Neurol, 2008. **7**(2): p. 136-44.
31. Merkies, I.S., et al., *Connecting impairment, disability, and handicap in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2003. **74**(1): p. 99-104.
32. Dalakas, M.C., *Pathogenesis and Treatment of Anti-MAG Neuropathy.* Curr Treat Options Neurol, 2010. **12**(2): p. 71-83.
33. Bland, J.M. and D.G. Altman, *Multiple significance tests: the Bonferroni method.* Bmj, 1995. **310**(6973): p. 170.

Chapter 5

Introducing the concept of minimum clinically important difference (MCID)

Chapter 5.1

MCID: general aspects

The concept of minimum clinically important difference (MCID): general aspects

One of the most important aspects of evaluating treatment effect is to define whether a change in clinical condition is relevant or not. Trial results are often based on statistical significant differences between the change scores of a treatment group and a control group. A difference may be statistically significant, in other words not based on chance or error alone, however whether this implies a clinically relevant change is questionable. Relevant change is captured in the terminology 'minimum clinically important difference (MCID)'.[1-3] The concept of MCID was defined by Jaeschke as being the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management.[2]

Although several parties will recognise the need and share an interest in determining MCID thresholds it is not a simple concept, nor simple to calculate. No fixed method to establish MCIDs is available.[4] In general, the available methods can be categorised as distribution-based or anchor-based. Distribution-based approaches are built upon measures of variablity of the obtained scores, like the standard deviation (SD), the standard error of measurement (SEM), the smallest detectable change (SDD) of effect size (ES). In contrast, anchor-based approaches compare the obtained scores with an external criterion, like a global assessment rating in which patient rate themselves as better, stable or worse.

The concept of MCID has hardly been addressed in immune-mediated neuropathy studies. In chapter 5.2 data of the ICE trial (placebo versus IVIg in CIDP) were used to illustrate various MCID methods to determine a responder, thereby shifting from statistical significant differences to clinically relevant differences. Chapter 5.3 highlights the use of variable individual standard errors as derived from Rasch-built interval outcome measures as possible MCID method for efficacy and maintenance trials in immune-mediated neuropathies.

References

1. Copay, A.G., et al., *Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales.* Spine J, 2008. 8(6): p. 968-74.
2. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status. Ascertaining the minimal clinically important difference.* Control Clin Trials, 1989. 10(4): p. 407-15.
3. Sloan, J.A., *Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions.* Copd, 2005. 2(1): p. 57-62.
4. Beaton, D.E., M. Boers, and G.A. Wells, *Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research.* Curr Opin Rheumatol, 2002. 14(2): p. 109-14.

Chapter 5.2

Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance

Ingemar S.J. Merkies[1,2], Sonja I. van Nes[2], Kim Hanna[3], Richard A.C. Hughes[4], Chunqin Deng[3]

[1]Department of Neurology, Spaarne Hospital, Hoofddorp, the Netherlands
[2]Department of Neurology, Erasmus Medical Centre, Rotterdam, the Netherlands
[3]Department of Clinical Development, Talecris Biotherapeutics, Inc, Research Triangle Park, North Carolina, USA
[4]Department of Clinical Neuroscience, King's College London, United Kingdom

J Neurol Neurosurg Psychiatry 2010 Nov;81(11):1194-9

Abstract

Background: The ICE trial demonstrated the efficacy of immune globulin intravenous (IGIV-C: Gamunex®) over placebo in chronic inflammatory demyelinating polyradiculoneuropathy (CIDP). However, improving the interpretability of the results by analysing the minimum clinically important difference (MCID) had not been considered.

Objectives: To identify MCID thresholds of various outcome measures using different methods and to test treatment differences (IGIV-C vs. placebo) using these thresholds.

Methods: One anchor-based (Short Form-36 question 2) and three distribution-based ($^1/_2$ standard deviation, 1 standard error of measurement, and effect size) techniques were employed to identify MCID cut-offs for various impairments (electromyographic parameters, Medical Research Council [MRC] sumscore, grip strength, Inflammatory Neuropathy Cause and Treatment [INCAT] sensory sumscore), disability (INCAT disability scale score, Rotterdam handicap scale [RHS] score), and quality of life (SF-36). IGIV-C or placebo was administered every 3 weeks for up to 24 weeks to 117 CIDP patients. Patients who did not improve by ≥1 point on the INCAT disability scale received alternate treatment. The proportion of patients with results exceeding identified MCID thresholds was compared.

Results: MCID cut-offs for outcomes were determined using each method. For the INCAT disability scale (primary ICE-trial outcome), all MCID methods identified significantly more responders with IGIV-C than placebo. Significant differences favouring IGIV-C were also demonstrated for various nerve conduction parameters, MRC sumscore, grip strength, RHS score, and SF-36 physical component summary score.

Conclusion: In addition to being statistically significant, all MCID analyses showed that CIDP improvements with IGIV-C are clinically meaningful. Consideration of MCID is recommended in future therapeutic trials.

Introduction

Chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) is characterised by a clinical course that may show either a continuous or stepwise progression or may have a more fluctuating pattern over months to years.[1, 2] Patients with CIDP often have a predominantly symmetrical distal weakness and sensory deficits that may lead to considerable long-term disability with decrement in health-related quality of life expectations.[1-4]

Recently, the IGIV-C CIDP efficacy trial (ICE trial), the largest randomised clinical trial of any agent for the treatment of CIDP ever performed, demonstrated the efficacy and safety of immune globulin intravenous, 10% caprylate/chromatography purified (IGIV-C).[4] Significant statistical differences in favour of IGIV-C were observed when comparing the results obtained with the Inflammatory Neuropathy Cause and Treatment (INCAT) disability scale as well as other outcome measures. However, despite these robust findings, no clinical trial in CIDP had ever examined whether the statistically significant results reflect clinically relevant changes for patients. Minimum clinically important difference (MCID) has been advanced as a concept that could enhance the interpretability of results of clinical studies and overcome the shortcomings of the 'statistically significant difference'.[5, 6] Jaeschke and associates first defined MCID as being 'the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management'.[7] Despite this seemingly clear definition, the MCID concept has been criticised because of difficulties with its calculation and because different methods produce different MCID values.[8] A number of techniques have been used to address these issues.[5, 6, 8]

The current study systematically addresses, for the first time in any neurological condition, the concept of MCID in a wide variety of outcome measures that capture disease progression in CIDP. Data were analysed from the ICE trial data to discover whether MCID techniques capture the efficacy of IGIV-C compared with placebo during up to 24 weeks of therapy.

Methods

Patients and study design

Complete details of the patient population and the randomised, double-blind, placebo-controlled, response-conditional study design and methodology are described in the publication of the primary efficacy analysis.[4] Eligible patients developed CIDP, progressive or relapsing motor and sensory dysfunction resulting from neuropathy over at least 2 months before study entry, and clinical disability as defined by an

overall INCAT disability score of 2-9.[9] Eligible patients were randomised to receive either IGIV-C (Gamunex®, Talecris Biotherapeutics, Inc, Research Triangle Park, North Carolina, USA) or placebo (0.1% albumin). Patients received a baseline loading dose of 2 g/kg and then a maintenance infusion of 1 g/kg every 3 weeks for up to 24 weeks. The selected outcome measures applied in the ICE study were based on various consensus meetings addressing standardisation of outcome measures in immune-mediated neuropathies; the WHO's International Classification of Functioning, Disability, and Health (ICF); and the quality of life concept.[10-13] All selected measures had demonstrated acceptable interpretability, good validity, reliability, and responsive values in patients with CIDP, and scales used in the ICE study were applied in a standardised fashion.[14] For the purpose of the current study, analyses were performed using the first-period data (up to 24 weeks) based on the intention-to-treat principle.

MCID determination

Because there is no global consensus on which technique is optimal for calculating MCID, this study employed four methods for quantitatively determining MCID, using either patient perception or mathematical modelling to calculate an MCID threshold value.[6, 8] One of the methods included in this study, an anchor-based method, compares changes in scale scores with those of another measurement or phenomenon that has known clinical relevance (e.g., patient perception of a clinical change after therapy). Three other techniques for calculating MCID, distribution-based methods, are determined from the statistical characteristics of the scores obtained (e.g., effect size).[5, 6, 8] The various threshold MCID values were then applied to assess the clinical importance of changes in outcome data from the first period of the ICE study. A goal of the study was to determine whether the selected MCID techniques would consistently rank all the scales used in favour of the IGIV-C group compared with the placebo group.

Anchor-based method

The global change of health assessment used the answers to Short Form-36 (Medical Outcomes Trust, Boston, MA, USA) question number 2 in which patients were asked to compare their global health rating at entry and at the end of the first period (week 24 receiving IGIV-C or placebo in a blinded fashion) versus 1 year ago.[15] This question is not used in scoring the domains and summary measure of the SF-36. This question asked patients to compare their current health to their health 1 year ago. Patient response options were 'much better', 'somewhat better', 'about the same', 'somewhat worse', or 'much better' now compared with 1 year ago. Various MCID techniques have been applied using the response options of the SF-36 question number 2.[16, 17] The MCID for the current study was based on a widely used approach and was determined as the mean score difference between the response options 'somewhat better now than 1 year ago' and 'about the same as 1 year ago'.[17, 18] Separately for both the IGIV-C and placebo groups, the patients were classified into two subgroups depending on whether their change in health score from baseline visit to week 24 was 1 (corresponding to relevant improvement) or 0 (corresponding to no improvement).

Distribution-based methods

The *one-half standard deviation (½ SD)* benchmark of an outcome measure has been proposed as the 'unifying theory' for assessing the MCID and has gained support in the literature.[6, 19-21] The SD for all outcome measures is based on their theoretical range. For example, the theoretical range for the grip strength, using the Vigorimeter (Martin, Tuttlingen, Germany), was based on available normative data.[22]

The *standard error of measurement (SEM)* was used to evaluate the intraindividual changes in selected outcome measures in both patient groups.[23] The SEM is equivalent to the SD of a scale multiplied by the square root of 1 minus its reliability coefficient (Cronbach's α). A change of 1 SEM has empirically been found to correspond with the original MCID criterion of an outcome measure.[23, 24] For each scale, the Cronbach α was calculated using only the placebo group with the repeat measurements at the baseline and at week 24 or the last measurement of the first period.

The *effect size* is the mean change in scores divided by the SD of the baseline scores. In the current study for the first period, the equation was: effect size = $(\mu_{24} - \mu_o)/SD\mu_o$, where μ_{24} = mean scale value at week 24 or the last visit of the first period and μ_o = mean scale value at baseline.[25] The effect size was calculated for both the IGIV-C and placebo groups. An effect size of 0.5 was considered an appropriate definition of MCID and was used in the current study.[6, 26]

Applying calculated MCID values

The MCID cut-off values were determined for each outcome measure applied in the ICE study based on all four MCID techniques. A patient was defined as a responder for a particular scale if the clinical condition of the patient improved enough to reach the MCID threshold for that scale. The percentage of patients in each treatment group was then calculated for those who reached the predefined MCID thresholds at the end of the first period, and the percentages were compared between the groups (Fisher exact two-tailed test). A p-value <0.05 was considered statistically significant. SAS version 8.2 was used for all statistical analyses (SAS institute, Cary, North Carolina, USA).

Results

During the first period of the ICE study, 117 individuals were randomised; 59 patients received IGIV-C and 58 patients received placebo.[4] Baseline values for the impairment measures (electromyographic parameters, Medical Research Council [MRC] sumscore, grip strength, and Inflammatory Neuropathy Cause and Treatment [INCAT] sensory sumscore), activity and participation scale scores (INCAT disability scale score and Rotterdam handicap scale [RHS] score), and quality of life component measures (SF-36 physical component score [PCS] and SF-36 mental component score [MCS]) were similar between the two treatment groups.[3,4,27]

MCID calculations

Determination of MCID was calculated using four methods, an anchor-based approach (SF-36, question 2) and three distribution-based approaches ($^1/_2$ SD, 1 SEM, and effect size). The calculated MCID cut-off scores of all measures for the anchor-based, $^1/_2$-SD, and 1 SEM methods are shown in table 1. To strive for uniformity, the mean MCID for the grip strength of dominant and non-dominant hands was calculated and applied to both hands. As indicated in the methods, the fourth technique of identifying an effect size of 0.5 as a cut-off value for MCID was also applied.

Applying MCID calculations to the ICE study

Anchor-based MCID method

All outcome measures except the MCS scores demonstrated higher percentages of improved values for the IGIV-C–treated patients compared with the placebo group when using the SF-36 question number 2 as the anchor (table 2). The INCAT disability score (the primary endpoint of the ICE study) showed a statistically significant difference in the percentage of patients reaching the MCID cut-off values in favour of the IGIV-C group. Significant differences in favour of IGIV-C were also observed for changes in average amplitude for all motor nerves, average conduction velocity for all motor nerves, percentage changes in conduction block, grip strength (non-dominant hand), MRC sumscore, and RHS score (table 2). A consistent trend in favour of IGIV-C was also observed for all outcome measures using the MCID cut-off values determined by $^1/_2$ SD and 1 SEM methods (table 2).

Table 1. Anchor-based*, $^1/_2$ SD, and 1 SEM distribution-based MCID techniques for the outcome measures applied in the ICE trial (first period of ICE study)

	Global health mean change from baseline using anchor-based approach*			½ SD		SEM	
Parameter	About the same† n=20	Somewhat better† n=25	MCID cut-off value	SD	MCID cut-off value	Cronbach's α	1 SEM as MCID
CMAP of most severely affected motor nerve, mV	0.54 ± 1.57	0.96 ± 1.92	**0.42**	1.73	**0.86**	0.77	**0.83**
CMAP for all motor nerves, mV	-0.14 ± 1.18	0.78 ± 0.85	**0.92**	2.30	**1.15**	0.90	**0.73**
Conduction velocity for all motor nerves, m/s	0.31 ± 4.95	2.52 ± 4.42	**2.21**	11.15	**5.58**	0.96	**2.23**
Conduction block reduction, %	0.79 ± 11.18	-2.28 ± 14.63	**3.07**	21.52	**10.76**	0.91	**6.46**
MRC sumscore	-0.3 ± 5.06	3.3 ± 3.71	**3.6**	7.06	**3.53**	0.92	**2.00**
Grip strength—dominant hand, kPa	-2.73 ± 20.64	14.24 ± 15.87	**14‡**	16	**8**	0.89	**7.78**
Grip strength—non-dominant hand, kPa	5.88 ± 14.39	16.85 ± 16.59	**14‡**	16	**8**	0.92	**6.78**
INCAT sensory sumscore	0.6 ± 4.0	-0.3 ± 3.5	**-0.9**	4.86	**2.43**	0.81	**2.12**
INCAT disability score	-0.4 ± 0.9	-1.0 ± 1.2	**-0.6**	1.44	**0.72**	0.80	**0.64**
Rotterdam handicap scale score	-0.6 ± 5.8	2.8 ± 4.9	**3.4**	6.82	**3.41**	0.82	**2.89**
SF-36 mental component score	-0.06 ± 10.15	4.07 ± 10.84	**4.13**	11.82	**5.91**	0.68	**6.69**
SF-36 physical component score	-0.98 ± 7.45	7.89 ± 7.34	**8.87**	8.28	**4.14**	0.62	**5.10**

Legend to table 1. *Global health change using the SF-36 question number 2.[15] †Data are mean ± SD for each scale (24-week values minus entry values). ‡Mean values for both hands were taken for uniformity. CIDP = chronic inflammatory demyelinating polyradiculoneuropathy; CMAP = compound muscle action potential; ICE = IGIV-C CIDP Efficacy trial; IGIV-C = immune globulin intravenous, 10% caprylate/chromatography purified; INCAT = Inflammatory Neuropathy Cause and Treatment; MCID = minimum clinically important difference; MRC = Medical Research Council; SD = standard deviation; SEM = standard error of measurement; SF-36 = Short Form-36 health survey

Table 2. Responder analyses defined by the MCID cut-off values for anchor-based, $^1/_2$-SD, and 1 SEM techniques for the IGIV-C and placebo groups of patients with CIDP (first period of the ICE study)

Parameter MCID	MCID cut-off	Responders IGIV-C: placebo,* n (%)	Percentage of responders, IGIV-C minus placebo (95% CI)	p-value†
CMAP of most severely affected motor nerve, mV				
Anchor based	0.42	23 (39.0): 21 (36.2)	2.8 (-16.5, 22.0)	0.849
$^1/_2$ SD	0.86	19 (32.2): 13 (22.4)	9.8 (-0.8, 27.5)	0.301
1 SEM	0.83	19 (32.2): 13 (22.4)	9.8 (-0.8, 27.5)	0.301
Average CMAP for all motor nerves, mV				
Anchor based	0.92	21 (35.6): 9 (15.5)	**20.1 (3.0, 37.2)**	**0.019**
$^1/_2$ SD	1.15	17 (28.8): 7 (12.1)	**16.7 (0.8. 32.7)**	**0.038**
1 SEM	0.73	26 (44.1): 12 (20.7)	**23.4 (5.3, 41.5)**	**0.010**
Average velocity for all motor nerves, m/s				
Anchor based	2.21	24 (40.7): 8 (13.8)	**26.9 (9.8, 44.0)**	**0.002**
$^1/_2$ SD	5.58	17 (28.8): 4 (6.9)	**21.9 (6.9, 36.9)**	**0.003**
1 SEM	2.23	24 (40.7): 8 (13.8)	**26.9 (9.8, 44.0)**	**0.002**
Conduction block, %				
Anchor based	3.07	30 (50.8): 17 (29.3)	**21.5 (2.5, 40.6)**	**0.024**
$^1/_2$ SD	10.76	17 (28.8): 5 (8.6)	**20.2 (4.9, 35.5)**	**0.008**
1 SEM	6.46	23 (39.0): 14 (24.1)	14.8 (-3.5, 33.2)	0.112
MRC sumscore				
Anchor based	3.6	25 (42.4): 6 (10.3)	**32.0 (15.5, 48.6)**	**<0.001**
$^1/_2$ SD	3.53	25 (42.4): 6 (10.3)	**32.0 (15.5, 48.6)**	**<0.001**
1 SEM	2	35 (59.3): 14 (24.1)	**35.2 (16.8, 53.6)**	**<0.001**
Grip strength-dominant hand, kPa‡				
Anchor based	14	21 (36.2): 12 (20.7)	15.5 (-2.4, 33.4)	0.100
$^1/_2$ SD	8	29 (50.0): 15 (25.9)	**24.1 (5.3, 43.0)**	**0.012**
1 SEM	7.78	29 (50.0): 15 (25.9)	**24.1 (5.3, 43.0)**	**0.012**
Grip strength-non-dominant hand, kPa‡				
Anchor based	14	23 (39.7): 11 (19.0)	**20.7 (2.8, 38.5)**	**0.025**
$^1/_2$ SD	8	30 (51.7): 20 (34.5)	17.2 (-2.2, 36.7)	0.091
1 SEM	6.78	31 (53.4): 22 (37.9)	15.5 (-4.1, 35.2)	0.136
INCAT sensory sumscore				
Anchor based	0.9	26 (44.1): 20 (34.5)	9.6 (-9.7, 28.9)	0.345
$^1/_2$ SD	2.43	18 (30.5): 9 (15.5)	15.0 (-1.7, 31.7)	0.078
1 SEM	2.12	18 (30.5): 9 (15.5)	15.0 (-1.7, 31.7)	0.078
INCAT disability score				
Anchor based	0.6	31 (52.5): 13 (22.4)	**30.1 (11.8, 48.5)**	**0.001**
$^1/_2$ SD	0.72	31 (52.5): 13 (22.4)	**30.1 (11.8, 48.5)**	**0.001**
1 SEM	0.64	31 (52.5): 13 (22.4)	**30.1 (11.8, 48.5)**	**0.001**
Rotterdam handicap scale score				
Anchor based	3.4	21 (35.6): 8 (13.8)	**21.8 (5.0, 38.6)**	**0.010**
$^1/_2$ SD	3.41	21 (35.6): 8 (13.8)	**21.8 (5.0, 38.6)**	**0.010**
1 SEM	2.89	27 (45.8): 11 (19.0)	**26.8 (8.9, 44.7)**	**0.003**
SF-36 mental component score				
Anchor based	4.13	18 (30.5): 19 (32.8)	-2.3 (-20.8, 16.3)	0.844
$^1/_2$ SD	5.91	16 (27.1): 17 (29.3)	-2.2 (-20.2, 15.8)	0.839
1 SEM	6.69	16 (27.1): 13 (22.4)	4.7 (-12.6, 22.0)	0.669
SF-36 physical component score				
Anchor based	8.87	18 (30.5): 11 (19.0)	11.5 (-5.7, 28.7)	0.199
$^1/_2$ SD	4.14	22 (37.3): 15 (25.9)	11.4 (-7.0, 29.8)	0.234
1 SEM	5.1	21 (35.6): 14 (24.1)	11.5 (-6.7, 29.6)	0.226

Legend to table 2. *59 and 58 patients received IGIV-C and placebo, respectively. †Difference in percentage of patients in both groups (IGIV-C and placebo) reaching the MCID cut-off values were examined using Fisher exact test. ‡Mean values for both hands were taken for uniformity. Values that are significantly different versus placebo are shown in bold text. CI = confidence interval; CIDP = chronic inflammatory demyelinating polyradiculoneuropathy; CMAP = compound muscle action potential; ICE = IGIV-C CIDP efficacy trial; IGIV-C = immune globulin intravenous, 10% caprylate/chromatography purified; INCAT = Inflammatory Neuropathy Cause and Treatment; MCID = minimum clinically important difference; MRC = Medical Research Council; SD = standard deviation; SEM = standard error of measurement, SF-36 = Short form-36 health survey.

Distribution-based MCID methods

In the $^1/_2$ SD method, significant differences were obtained in favour of the IGIV-C group for the nerve conduction parameters, improvement in the average CMAP amplitude for all motor nerves, average conduction velocity for all motor nerves, and reduction in percentage of conduction block (table 2). More patients in the IGIV-C group reached the MCID cut-off values for the MRC sumscore and dominant hand grip strength ($p \leq 0.012$). Applying the MCID cut-off value obtained for the INCAT disability scale using the $^1/_2$ SD method demonstrated that a substantially higher percentage of patients in the IGIV-C group reached the MCID threshold (52.2% vs.

22.4% in IGIV-C and placebo groups, respectively; p=0.001). A significant difference in favour of the IGIV-C group was also seen for the MCID findings for RHS score.

An almost similar pattern was seen using the 1 SEM method to determine the MCID thresholds for the applied outcome measures (table 2). A substantial significant difference in favour of the IGIV-C group was observed for improvement in average amplitude for all motor nerves, improvement in average conduction velocity for all motor nerves, MRC sumscore, grip strength (dominant hand), INCAT disability score, and RHS score. For the INCAT sensory score and the component quality of life measures, the differences were not statistically significant.

For the fourth technique, using effect size to determine MCID, the effect size values were consistently higher in the IGIV-C group than with the scores in the placebo group (table 3). None of the applied outcome measures in the placebo group reached the MCID cut-off value of 0.5. In the IGIV-C group, the MCID cut-off value was obtained for INCAT disability score, grip strength, and PCS. The INCAT disability score had the highest effect size. The effect size for MRC sumscore, RHS score, and changes in amplitude for most severely affected motor nerve nearly reached the cut-off score of 0.5.

Table 3. Responder analyses comparing effect size scores in the IGIV-C and placebo groups of patients with CIDP (first period of the ICE study)

Parameter	Effect size*	
	IGIV-C group n=59	Placebo group n=58
CMAP of most severely affected motor nerve, mV	0.493	0.235
Averaged CMAP from all motor nerves, mV	0.343	0.065
Averaged velocity from all motor nerves, m/s	0.245	0.041
Conduction block, %	0.236	0.064
MRC sumscore	0.474	0.026
Grip strength-dominant hand, kPa	**0.557**	0.066
Grip strength-non-dominant hand, kPa	**0.532**	0.187
INCAT sensory sumscore	0.237	0.050
INCAT disability score	**0.754**	0.226
Rotterdam handicap scale score	0.458	0.071
SF-36 mental component score	0.111	0.215
SF-36 physical component score	**0.719**	0.170

Legend to table 3. *Effect size was defined as: $(\mu_{24} - \mu_0)/SD\mu_0$; μ_{24} = mean scale value of the longitudinally examined group at week 24; μ_0 = mean scale value at week = 0 [entry].[25] An effect size of 0.5 was defined as the MCID.[6]Values that were significantly ($p<0.05$) different versus placebo are shown in bold text. CIDP = chronic inflammatory demyelinating polyradiculoneuropathy; CMAP = compound muscle action potential; ICE = IGIV-C CIDP efficacy trial; IGIV-C = immune globulin intravenous, 10% caprylate/chromatography purified; INCAT = Inflammatory Neuropathy Cause and Treatment; MCID = minimum clinically important difference; MRC = Medical Research Council; SF-36 = Short Form-36 health survey.

Discussion

The current study is the first in the field of peripheral nerve treatment trials to use MCID methods to evaluate treatment response. It considered measures of impairment, activity and participation, and quality of life and used both anchor-based and distribution-based methods on the data of the ICE trial.[4] The four methods evaluated resulted in varied MCID cut-off values, an observation that has been reported in other studies.[5, 6, 8, 28] However, in the current study, the threshold variation among the four MCID methods was smaller for the INCAT disability score and RHS score than for other outcome measures. Thus, for these scales, the MCID threshold is likely to be well defined and corroborated by all four methods used.

The INCAT disability score consistently demonstrated a higher percentage of responders reaching the calculated MCID threshold in the IGIV-C group compared with the placebo group. The differences for the INCAT disability score were always statistically significant regardless of the MCID method used, thereby neutralising any debate about the optimal method for the MCID calculation in this study. Using the $^1/_2$ SD as the unified theory MCID technique, a total of 52.2% of the IGIV-C group reached the MCID threshold for the INCAT disability score compared with 22.4% in the placebo arm in the first period of treatment, which is almost identical to the calculation of the number improving by ≥1 point, the primary outcome in the trial.[4] Similar percentage scores were found when applying the 1 SEM distribution and SF-36 question number 2 as methods for assessing MCID. Using the effect size method, the INCAT disability scale demonstrated the highest value in the IGIV-C group and can therefore be considered the most sensitive to capture responsiveness when compared with other outcome measures applied.

A systematic trend in favour of the IGIV-C group was also seen when evaluating the proportion of responders demonstrating changes greater than the MCID thresholds for the other outcome measures. These differences were robust in both the anchor-based and distribution-based techniques, with significant changes seen for various nerve conduction study parameters, MRC sumscore, grip strength, RHS score, and effect size for the quality of life PCS.

The anchor-based method using the SF-36 question number 2 to identify changes in global health has been criticised because patient answers may be influenced by health issues unrelated to the illness examined. The retrospective rating using SF-36 question number 2 with an extended period of time of 1 year is probably susceptible to a number of known recall biases.[29] In light of these biases, global ratings may not give an accurate picture of an individual's 'true' change. Also, global ratings may only explain parts of the health-related quality of life findings.[18] However, despite these reported shortcomings, the results in the current study with the SF-36 question number 2 as the anchor were comparable to the results seen in the various distribution-based methods, despite the differences in their cut-off

values for the various outcome measures. Similar comparable clinically meaningful changes were also reported in patients with lung cancer using anchor-based and distribution-based MCID methods.[30]

Another important issue is whether the baseline level of functioning, disability, or health status affects MCID thresholds. Do more severely affected patients with CIDP require a greater change from baseline to be considered clinically meaningful than patients less severely affected?[28, 31] In a study of patients with low back pain, a greater MCID change was required in patients with initial higher pain intensity than those with less pain.[32] According to modern clinimetric approaches, clinically meaningful changes may vary across the theoretical range of an outcome measure.[33] Another important issue is whether to use a combination of an anchor-based and a distribution-based method to ascertain improvement in patients with CIDP.[30, 34] To compensate for the differences seen in cut-off values for some outcome measures in the current paper and in the light of lack of consensus on this topic, we propose for future trials that patients with CIDP should be considered *improved* only when they meet 1 anchor-based plus 1 distribution-based MCID technique for change (the so-called 'combined MCID robustness approach') as has been addressed by others.[30, 34, 35] The combination of SF-36 question number 2 (anchor-based form) plus the unified distribution-based method of $^1/_2$ SD is suggested in the absence of any current consensus.

There are some methodological limitations to the current study that should be addressed. One overall limitation of using MCID techniques is the applicability to nonlinear scales. Most of the outcome measures evaluated in the current study were nonlinear, except for grip strength and neurophysiologic data. Therefore, the analyses in the current study should be interpreted with some caution because it is conceivable that a calculated MCID may vary over the range of a nonlinear scale. Also, the MCID approach in the current study should not be simply extrapolated to other illnesses. The current MCID results could, however, be useful in trials with a similar patient population. The results also serve as a first step and framework for defining appropriate MCIDs in other chronic neurologic conditions. Importantly, the *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study group has developed a linearly weighted overall disability scale for immune-mediated peripheral neuropathies, aiming to bypass the above postulated shortcomings and to increase the future applicability of MCID techniques in these conditions.[36]

Rheumatologists have used consensus meetings to determine the optimal outcome measures and methods of analysis, and neurologists have begun to follow a similar approach.[10, 11, 31, 37-39] Additional consensus meetings should be organised to help researchers in the field to determine whether MCID should be based on the patients' or the clinicians' points of view, whether more emphasis should be put on objective measures rather than subjective measures, whether future studies should be based on a desirable amount of change rather than a minimal measurable

change, and whether inferences about MCID are made with respect to individuals or groups of patients.[5-7, 18, 28, 31, 37-39]

In conclusion, the current paper shows how MCID techniques for a variety of outcome measures can be used to compare disease progression in a CIDP treatment trial. The MCID approach adds a dimension of clinical relevance to statistical significance and to the conclusions about the efficacy of IGIV-C. The concept of MCID is an important contribution to the interpretation of clinical trials and deserves further consideration in this and other fields of neurology.

References

1. Koller, H., et al., *Chronic inflammatory demyelinating polyneuropathy.* N Engl J Med, 2005. **352**(13): p. 1343-56.
2. Barohn, R.J., et al., *Chronic inflammatory demyelinating polyradiculoneuropathy. Clinical characteristics, course, and recommendations for diagnostic criteria.* Arch Neurol, 1989. **46**(8): p. 878-84.
3. Merkies, I.S., et al., *Health-related quality of life improvements in CIDP with immune globulin IV 10%: the ICE Study.* Neurology, 2009. **72**(15): p. 1337-44.
4. Hughes, R.A., et al., *Intravenous immune globulin (10% caprylate-chromatography purified) for the treatment of chronic inflammatory demyelinating polyradiculoneuropathy (ICE study): a randomised placebo-controlled trial.* Lancet Neurol, 2008. **7**(2): p. 136-44.
5. Copay, A.G., et al., *Understanding the minimum clinically important difference: a review of concepts and methods.* Spine J, 2007. **7**(5): p. 541-6.
6. Sloan, J.A., et al., *Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials.* Drug Inf J. , 2003. **37**: p. 23–31.
7. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status. Ascertaining the minimal clinically important difference.* Control Clin Trials, 1989. **10**(4): p. 407-15.
8. Beaton, D.E., M. Boers, and G.A. Wells, *Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research.* Curr Opin Rheumatol, 2002. **14**(2): p. 109-14.
9. Hughes, R., et al., *Randomized controlled trial of intravenous immunoglobulin versus oral prednisolone in chronic inflammatory demyelinating polyradiculoneuropathy.* Ann Neurol, 2001. **50**(2): p. 195-201.
10. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
11. Lunn, M.P., et al., *151st ENMC international workshop: Inflammatory Neuropathy Consortium 13th-15th April 2007, Schiphol, The Netherlands.* Neuromuscul Disord, 2008. **18**(1): p. 85-9.
12. World Health Organisation, *International classification of impairments, disabilities, and handicaps*. 2001: Geneva.
13. Aaronson, N.K., *Quality of life: what is it? How should it be measured?* Oncology (Williston Park), 1988. **2**(5): p. 69-76, 64.
14. Merkies, I.S.J., *Evaluation of scales and measurement instruments in immune-mediated polyneuropathies*, in *Department of Neurology*. 2001. ISBN: 90-9014393-9, Erasmus Medical Centre: Rotterdam, The Netherlands.
15. Ware, J.E., Jr, M. Kosinski, and B. Gandek, *SF-36® Health Survey. Manual and Interpretation Guide*. 2000, Lincoln, RI: QualityMetric Incorporated.
16. Chiou, C.F., et al., *Development and validation of the revised Cedars-Sinai health-related quality of life for rheumatoid arthritis instrument.* Arthritis Rheum, 2006. **55**(6): p. 856-63.
17. Angst, F., A. Aeschlimann, and G. Stucki, *Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities.* Arthritis Rheum, 2001. **45**(4): p. 384-91.
18. Copay, A.G., et al., *Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales.* Spine J, 2008. **8**(6): p. 968-74.
19. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation.* Med Care, 2003. **41**(5): p. 582-92.
20. Juniper, E.F., et al., *Determining a minimal important change in a disease-specific Quality of Life Questionnaire.* J Clin Epidemiol, 1994. **47**(1): p. 81-7.
21. Cella, D.F., *Quality of life outcomes: measurement and validation.* Oncology (Williston Park), 1996. **10**(11 Suppl): p. 233-46.
22. Merkies, I.S., et al., *Assessing grip strength in healthy individuals and patients with immune-mediated polyneuropathies.* Muscle Nerve, 2000. **23**(9): p. 1393-401.

23. Wyrwich, K.W., W.M. Tierney, and F.D. Wolinsky, *Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life.* J Clin Epidemiol, 1999. **52**(9): p. 861-73.
24. Wyrwich, K.W., *Minimal important difference thresholds and the standard error of measurement: is there a connection?* J Biopharm Stat, 2004. **14**(1): p. 97-110.
25. Kazis, L.E., J.J. Anderson, and R.F. Meenan, *Effect sizes for interpreting changes in health status.* Med Care, 1989. **27**(3 Suppl): p. S178-89.
26. Cohen, J., *Statistical power analysis for the behavioral sciences*. 1988, Hillsdale, NY: Lawrence Erlbaum Associates, Inc.
27. Bril, V., et al., *Electrophysiology in chronic inflammatory demyelinating polyneuropathy with IGIV.* Muscle Nerve, 2009. **39**(4): p. 448-455.
28. Osoba, D., et al., *Interpreting the significance of changes in health-related quality of life scores.* J Clin Oncol, 1998. **16**(1): p. 139-44.
29. Schwartz, N., *Autobiographical memory and the validity of retrospective reports*. 1994, New York, NY: Springer-Verlag.
30. Cella, D., et al., *What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592.* J Clin Epidemiol, 2002. **55**(3): p. 285-95.
31. Crosby, R.D., R.L. Kolotkin, and G.R. Williams, *Defining clinically meaningful change in health-related quality of life.* J Clin Epidemiol, 2003. **56**(5): p. 395-407.
32. Stratford, P.W., et al., *Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1.* Phys Ther, 1998. **78**(11): p. 1186-96.
33. Lai, J.-S. and D.T. Eton, *Clinically meaningful gaps.* Rasch Measurement Transactions, 2002. **15**: p. 850.
34. Kolotkin, R.L., R.D. Crosby, and G.R. Williams, *Integrating anchor-based and distribution-based methods to determine clinically meaningful change in obesity-specific quality of life.* Qual Life Res, 2002. **11**: p. 670.
35. Jacobson, N.S. and P. Truax, *Clinical significance: a statistical approach to defining meaningful change in psychotherapy research.* J Consult Clin Psychol, 1991. **59**(1): p. 12-9.
36. van Nes, S., et al., *A rasch-built, linearly-weighted overall disability sumscore for immune-mediated peripheral neuropathies. Presented at the 2009 Meeting of the Peripheral Nerve Society, July 4-8, 2009.* 2009.
37. Beaton, D.E., et al., *Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference.* J Rheumatol, 2001. **28**(2): p. 400-5.
38. Wells, G., et al., *MCID/Low Disease Activity State Workshop: summary, recommendations, and research agenda.* J Rheumatol, 2003. **30**(5): p. 1115-8.
39. Guyatt, G.H., et al., *Methods to explain the clinical significance of health status measures.* Mayo Clin Proc, 2002. **77**(4): p. 371-83.

Chapter 5.3

Defining a responder: a dynamic changing pattern in immune-mediated neuropathies through Rasch analyses

Els K. Vanhoutte[1], Sonja I. van Nes[2], Pieter A. van Doorn[2], Kenneth C. Gorson[3], David R. Cornblath4, Catharina G. Faber[1], Ingemar S.J. Merkies[1,2,5], on behalf of the PeriNomS study group

[1]Department of Neurology, University medical centre Maastricht, Maastricht, the Netherlands
[2]Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands
[3]Department of Neurology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, USA
[4]Department of Neurology, Johns Hopkins Hospital and School of Medicine, Baltimore, USA
[5]Department of neurology, Spaarne Hospital, Hoofddorp, the Netherlands

Manuscript in preparation

Abstract

Background: Often, studies report as the primary outcome whether a statistically significant difference between a placebo and intervention group was present, without taking the concept of a minimum clinically important difference (MCID) into account. When the MCID is considered to define a responder, usually a fixed cut-off value, like a fixed estimate of the standard error (SE), is used. However, Rasch analysis has demonstrated that the individual SE varies, with higher SE at the extremes than at the centre of the scale of an interval outcome measure. Therefore, defining a responder using the concept of the MCID based on the individual SE should also have a dynamic pattern

Objective: To examine the dynamic pattern of being a responder through the concept of minimum clinically important difference using changing SE values (MCID-SE) in patients with Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP), and multifocal motor neuropathy (MMN).

Methods: Ninety-two newly diagnosed patients (GBS: 36, CIDP: 35, MGUSP: 11, MMN: 10) were serially examined during 1 year of follow-up. MCID-SE was determined in each patient for each outcome measure (GBS/CIDP/MGUSP: Rasch-built MRC sumscore, Rasch-built modified INCAT sensory scale, Rasch-built overall disability scale (R-ODS); MMN: Rasch-built modified MRC sumscore, R-ODS for MMN). Since GBS and CIDP are characterised by an acute or subacute onset, as opposed to the more indolent course of MGUSP and MMN, and since there is no consensus on the magnitude of MCID-SE, different cut-off values were examined (1x and 1.96x SE).

Results: In all four diseases, the changing SE values throughout all outcome measures captured the direction of changes (improvement, stable, deterioration). The number of patients being a responder varied depending on the cut-offs and type of illness, and were high for the R-ODS.

Conclusion: The concept of MCID-SE using variable SE values for several interval outcome measures demonstrated a dynamic pattern of defining a responder in patients with immune-mediated neuropathies. Future intervention studies in these conditions should thus consider using the concept of MCID-SE to capture the dynamics of defining a responder.

Introduction

Clinical interventional trials performed in patients with immune-mediated neuropathies like Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), monoclonal gammopathy of undetermined significance related polyneuropathy (MGUSP) and multifocal motor neuropathy (MMN) have generally used traditional ordinal based outcome measures in the absence of alternative linearly constructed measures.[1-6] These studies have generally used a fixed cut-off score to determine responders among patients receiving a particular intervention. The approach of a 'fixed cut-off' to define a responder also implied a fixed estimate of the standard error (SE) across the whole range (continuum) of the scales used.[7] Based on the methodology of a fixed standard error, the efficacy of intravenous immunoglobulin (IVIg) over placebo was demonstrated in patients with CIDP through the concept of minimum clinically important difference (MCID) (see also chapter 5.2).[8] A comparable approach has also been described by others.[9-11] However, modern methods like Rasch analysis have shown variable SE values related to different patient locations on the metric applied, thus suggesting a dynamic changing pattern of defining a responder as opposed to the traditional approach for assessing outcome.[12-15]

No interventional study in patients with GBS, CIDP, MGUSP or MMN has examined the varying SE along the continuum of outcome measures applied. The purpose of the current paper is to illustrate the variable SE across various impairment and disability measures applied in these conditions, after converting these to interval measures with a linear construct using the Rasch method.[11, 12] The dynamics of defining a responder are demonstrated through the concept of minimum clinically important difference using the variable SE values (MCID-SE) in serially examined patients receiving interventional therapy.

Methods

Patients and examination schedule

Patients were recruited between July 2008 and November 2010 as part of the on-going large, international, multi-centre collaborative study, the *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study that aims to improve and standardise the clinical assessment of patients with immune-mediated neuropathies through a comprehensive clinimetric approach. A total of 92 newly diagnosed patients with GBS (36), CIDP (35), MGUSP (11), and MMN (10) were recruited. Patients with GBS and CIDP were examined at entry, 1, 3, 6, and 12 months and patients with MMN and MGUSP at entry, 3, and 12 months of follow-up. All patients with GBS were treated at entry; 23 with IVIg, 6 with plasma

exchange, 5 with IVIg combined with prednisolone and 2 with IVIg after plasma exchange. All but one of the CIDP patients have been treated during follow-up, most of them (18) with IVIg, the others receiving corticosteroids (6), plasma exchange (1), immunosuppressive drugs (3) or a combination of IVIg with corticosteroids and/or plasma exchange (6). Of the 11 MGUSP patients, 5 have been treated (2 with IVIg, 3 with Rituximab). All MMN patients except 1 were treated with IVIg; in one patient IVIg was combined with cyclophosphamide interval therapy.

All patients with GBS, CIDP, and MMN met the international criteria for their illness.[16-18] The diagnosis of MGUSP was established after excluding all other possible causes for gammopathy and polyneuropathy.[19]

Standard protocol approvals, registrations, and patient consents

The local medical ethics committee in each participating centre approved the protocol. Written informed consent was obtained from all participants.

Outcome measures applied

The selection of outcome measures was based on recommendations given at a workshop on outcome measures in immune-mediated neuropathies.[20] The following impairment and disability measures were selected for the patients with GBS, CIDP, and MGUSP:

- The *MRC sumscore* comprises the following six muscle pairs: upper arm abductors, elbow flexors, wrist extensors, hip flexors, knee extensors, and foot dorsal flexors.[21] These muscles were examined and scored using the Medical Research Council grades that range from 0 (paralysis) to 5 (normal strength).[22]
- The *modified INCAT sensory sumscore (mISS)* embraces pinprick, vibration, light touch, and joint position senses assessed separately in the arms and legs plus the two-point discrimination only assessed at the index finger (using a sliding aesthesiometer with curved sharp tips).[23] Except for the two-point discrimination, all sensory qualities were scored separately using response options that range from 0 (no sensory deficit at the index finger or hallux) to 4 (highest sensory deficit at acromioclavicular joint or anterior superior iliac spine). Two-point discrimination was scored as normal (grade 0) or disturbed (grade 1) based on age-dependent normative values.[24]
- The *Rasch-built overall disability scale (R-ODS)* consists of 24 daily and social activity items that are scored as 0 ('impossible to perform'), 1 ('performed with difficulty'), or 2 ('easily performed, without difficulty').[25] This outcome measure at the interval level was specifically designed to capture activity limitations and participation restrictions in patients with GBS, CIDP, and MGUSP.
- For the patients with MMN, the *modified MRC sumscore* was assessed of the following muscle pairs: shoulder abductors, elbow flexors, elbow extensors, wrist extensors, wrist flexors, finger extensors, fingers spreaders, finger flexors, thumb adductor, thumb adductor, and thumb opponent. Also, a 25- items MMN specific R-ODS (*R-ODS-MMN*) was applied.

Study procedures

All selected measures were applied according to the previously published requirements.[25, 26] All participating centres as part of the ***P*eri*N*om*S*** study (n=26) were trained in 2007 (by IM and SvN) at the Peripheral Nerve Society meeting (Utah, USA) aiming to standardise the assessment procedures for all scales as part of this study. In addition, participants received a research manual that included a thorough description and pictures illustrating how to assess the various sensory modalities for the MRC sumscore and the mISS. Patients also received standardised instructions for the completion of their illness-specific R-ODS.

Rasch analyses and statistical aspects

Creating interval measures through Rasch

The Rasch methodology has been described thoroughly elsewhere, also specifically for neurologists (see also chapter 3.1).[8, 12, 13] The collected data for all selected outcome measures were subjected to Rasch analyses using the software Rasch unidimensional measurement models (RUMM).[27] Data of MMN patients were collected and analysed separately, because all measures applied in MMN are illness-specific. The RUMM software enables the transformation of ordinal obtained scores into interval scores, and provides an estimate of each patient location with a corresponding SE on the same logit (log-odds unit) scale.

Determining MCID-SE based on the clinical picture

Various methods have been proposed to capture clinically relevant changes in patients receiving interventional therapies.[28-30] However, there is no consensus on how to quantify the importance and magnitude of such changes. Also no study in immune-mediated neuropathies has considered the possibility that the magnitude of clinical relevant change can vary, dependent on the clinical picture of the diagnosis and variable SE corresponding to the ability of the patient. Deficit in GBS and CIDP is characterised by an acute or subacute onset, mainly driven by motor dysfunction and less by sensory disturbances.[31] Therefore, it is conceivable to hypothesise that in GBS and CIDP motor changes as assessed with the MRC sumscore will be of a higher magnitude than sensory changes using the mISS. In contrast, the clinical picture of MGUSP and MMN tends to be more indolent with less vigorous changes. MMN generally presents with only motor deficits, whereas in MGUSP, sensory disturbances are more prominent than weakness. Clinical changes in MGUSP and MMN will probably be less pronounced.

Responsiveness at the individual person level

Referred to as minimum clinically important difference-standard error (MCID-SE) score based on the previously described significant change (SigChange).[14] In brief, MCID-SE was calculated by computing for each participating patient at each assessment and for each scale applied:

- their own change (person location at *i* month - person location at entry; where *i* = 1, 3, 6 or 12 months of follow-up in patients with GBS and CIDP or 3 or 12 months in patients with MGUSP and MMN),
- the corresponding SE of difference related to their individual change (SE_{diff} = square-root (SE_{entry}^2+$SE_{i\ month}^2$)), and
- the final MCID-SE calculations by dividing the individual change scores by corresponding SE_{diff} (MCID-SE = (person location at *i* month - person location at entry)/SE_{diff}).[13]

Since there is no consensus on the magnitude of MCID-SE, we have decided to perform two separate classifications using cut-off values for the MCID-SE scores at 1x SE or 1.96x SE for each outcome measure applied in the various illnesses. The degree of uncertainty associated with the ability estimate is expressed by the standard error, and can be seen as an estimate of the standard deviation of the ability of a patient. Thus, using 1 SE gives 68% certainty that the ability is in this range; using 1.96 SE gives 95% confidence.

Five different subgroups were created, based on the MCID-SE scores:

- subgroup 1 (clinically important improvement): MCID-SE ≥ 1 (*or* 1.96)
- subgroup 2 (clinically unimportant improvement): 0 < MCID-SE < 1 (*or* 1.96)
- subgroup 3 (no change): MCID-SE = 0
- subgroup 4 (clinically unimportant deterioration): -1 (*or* -1.96) < MCID-SE < 0
- subgroup 5 (clinically important deterioration): MCID-SE ≤ -1 (*or* -1.96)

Since a reduction in mISS score reflects improvement, the MCID-SE cut-offs were reversed for the mISS scale (subgroup 1 defined as clinically important improvement: MCID-SE ≤ -1 (*or* -1.96); subgroup 2 defined as clinically unimportant improvement: -1 (*or* -1.96) < MCID-SE < 0 ; subgroup 3 defined as no change: MCID-SE = 0; subgroup 4 defined as clinically unimportant deterioration: 0 < MCID-SE < 1 (*or* 1.96), and subgroup 5 defined as clinically important deterioration: MCID-SE ≥ 1 (*or* 1.96).

Software

Rasch analyses were performed with the partial credit model as default (RUMM2030). Graphs were constructed using Stata 11.0 for Windows XP.

Results

General aspects study population

The general characteristics of participants are described in table 1.

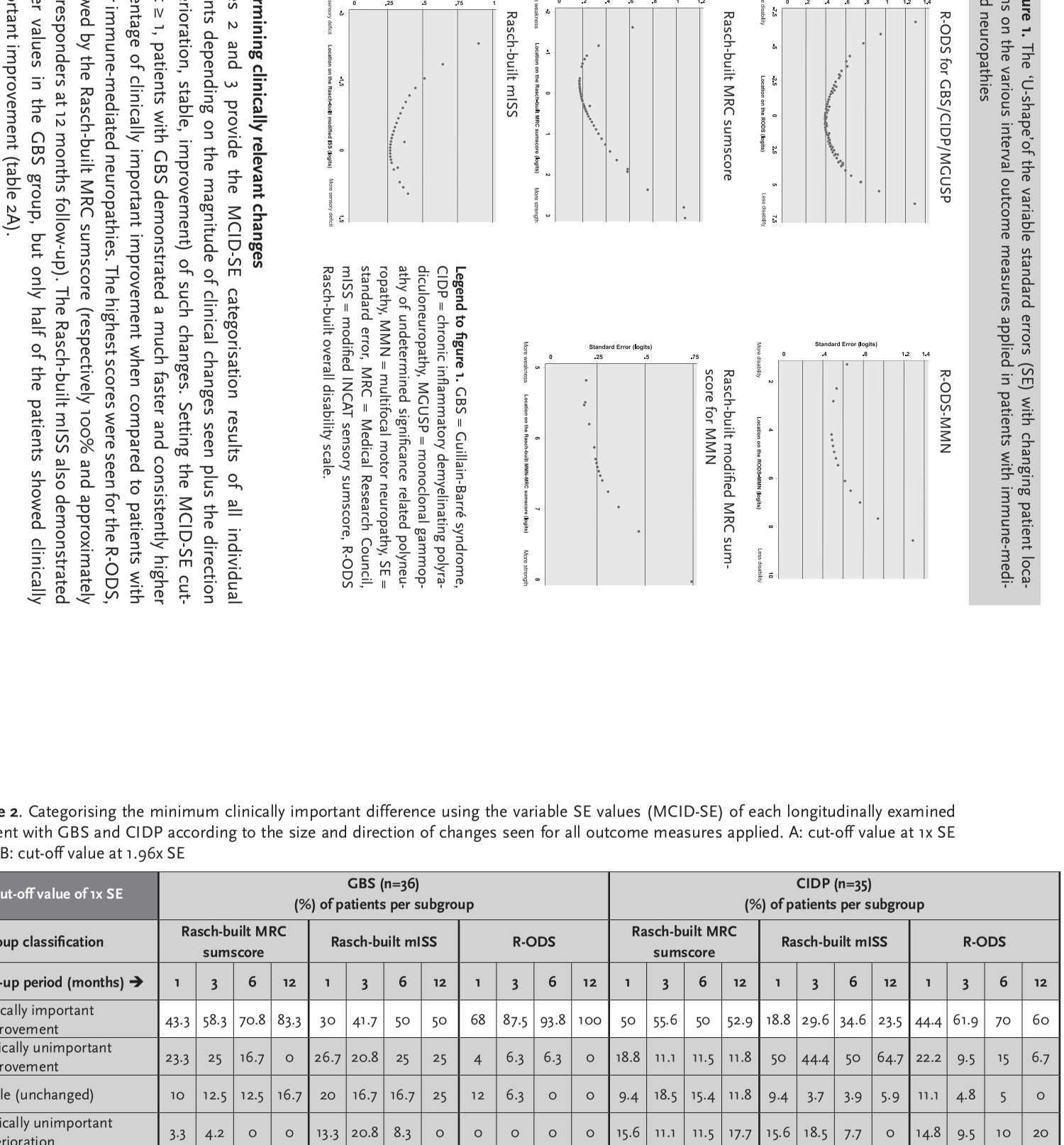
Table 1. General characteristics of the participants

	Total	GBS	CIDP	MGUSP	MMN
Number of patients	92	36	35	11	10
Age (years), mean (SD), range		55.9 (19.2), 19-90	53.3 (14.0), 18-74	60.7 (12.6), 45-85	46.5 (9.2), 32-60
Gender, n (%)					
Male	65 (71)	25 (69)	26 (74)	6 (55)	8 (80)
Female	27 (29)	11 (31)	9 (26)	5 (46)	2 (20)
Country of assessment, n (%)					
Netherlands	40 (43)	19 (53)	15 (43)	3 (27)	3 (30)
USA	22 (24)	6 (17)	9 (25)	3 (27)	4 (40)
Italy	13 (14)	6 (17)	3 (9)	2 (18)	2 (20)
Canada	4 (4)	2 (6)	2 (6)	-	-
Belgium	5 (5)	1 (3)	2 (6)	2 (18)	-
France	3 (3)	-	1 (3)	1 (9)	1 (10)
UK	2 (2)	1 (3)	1 (3)	-	-
Spain	1 (1)	-	1 (3)	-	-
Brazil	2 (2)	1 (3)	1 (3)	-	-

Variable SE with changing patient locations

The mISS and the MRC sumscore were subjected to RUMM software, hereby creating interval measures with corresponding locations and SE for each individual. Figure 1 illustrates the variable SE with changing patient locations on the Rasch-built outcome measures. As can be seen, a somewhat 'U-shape' dot-pattern is shown, with lower SE scores in the mid-section and increasing SE values towards the extremes of patients'ability levels in all metrics applied. The highest variation in SE was seen in the Rasch-built MRC sumscore for patients with GBS, CIDP, and MGUSP (fivefold). The SE varied twofold in the Rasch-built modified MRC sumscore for patients with MMN, threefold in the Rasch-built mISS, in the R-ODS for GBS, CIDP, and MGUSP patients, and in the R-ODS-MMN (figure 1).

Figure 1. The 'U-shape' of the variable standard errors (SE) with changing patient locations on the various interval outcome measures applied in patients with immune-mediated neuropathies



Legend to figure 1. GBS = Guillain-Barré syndrome, CIDP = chronic inflammatory demyelinating polyradiculoneuropathy, MGUSP = monoclonal gammopathy of undetermined significance related polyneuropathy, MMN = multifocal motor neuropathy, SE = standard error, MRC = Medical Research Council, mISS = modified INCAT sensory sumscore, R-ODS Rasch-built overall disability scale.

Determining clinically relevant changes

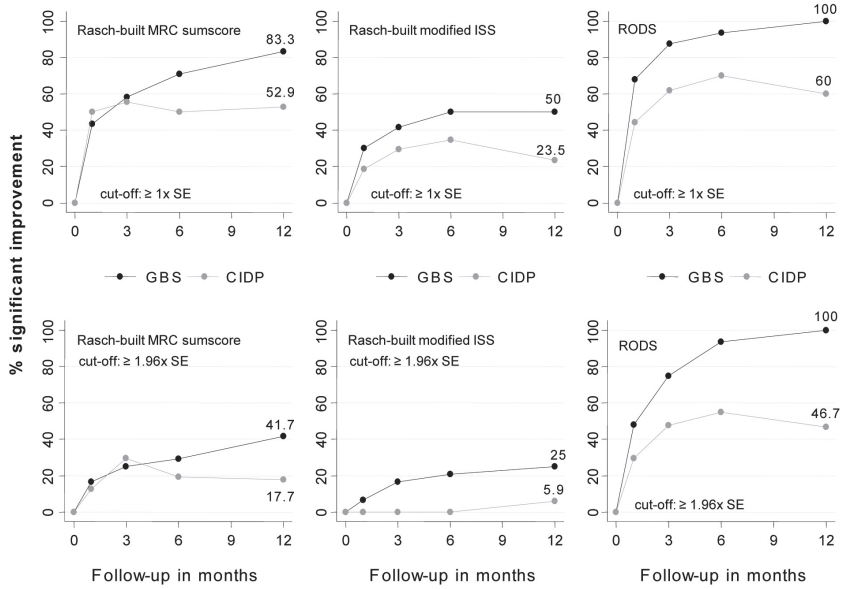
Tables 2 and 3 provide the MCID-SE categorisation results of all individual patients depending on the magnitude of clinical changes seen plus the direction (deterioration, stable, improvement) of such changes. Setting the MCID-SE cut-off at ≥ 1, patients with GBS demonstrated a much faster and consistently higher percentage of clinically important improvement when compared to patients with other immune-mediated neuropathies. The highest scores were seen for the R-ODS, followed by the Rasch-built MRC sumscore (respectively 100% and approximately 83% responders at 12 months follow-up). The Rasch-built mISS also demonstrated higher values in the GBS group, but only half of the patients showed clinically important improvement (table 2A).

Table 2. Categorising the minimum clinically important difference using the variable SE values (MCID-SE) of each longitudinally examined patient with GBS and CIDP according to the size and direction of changes seen for all outcome measures applied. A: cut-off value at 1x SE and B: cut-off value at 1.96x SE

A: Cut-off value of 1x SE	GBS (n=36) (%) of patients per subgroup												CIDP (n=35) (%) of patients per subgroup											
Subgroup classification	**Rasch-built MRC sumscore**				**Rasch-built mISS**				**R-ODS**				**Rasch-built MRC sumscore**				**Rasch-built mISS**				**R-ODS**			
Follow-up period (months) ➔	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**
1: Clinically important improvement	43.3	58.3	70.8	83.3	30	41.7	50	50	68	87.5	93.8	100	50	55.6	50	52.9	18.8	29.6	34.6	23.5	44.4	61.9	70	60
2: Clinically unimportant improvement	23.3	25	16.7	0	26.7	20.8	25	25	4	6.3	6.3	0	18.8	11.1	11.5	11.8	50	44.4	50	64.7	22.2	9.5	15	6.7
3: Stable (unchanged)	10	12.5	12.5	16.7	20	16.7	16.7	25	12	6.3	0	0	9.4	18.5	15.4	11.8	9.4	3.7	3.9	5.9	11.1	4.8	5	0
4: Clinically unimportant deterioration	3.3	4.2	0	0	13.3	20.8	8.3	0	0	0	0	0	15.6	11.1	11.5	17.7	15.6	18.5	7.7	0	14.8	9.5	10	20
5: Clinically important deterioration	20	0	0	0	10	0	0	0	16	0	0	0	6.3	3.7	11.5	5.9	6.3	3.7	3.9	5.9	7.4	14.3	0	13.3
B: Cut-off value of 1.96x SE	**GBS (n=36) (%) of patients per subgroup**												**CIDP (n=35) (%) of patients per subgroup**											
Follow-up period (months) ➔	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**
1: Clinically important improvement	16.7	25	29.2	41.7	6.67	16.7	20.8	25	48	75	93.8	100	12.5	29.7	19.2	17.7	0	0	0	5.88	29.6	47.6	55	46.7
2: Clinically unimportant improvement	50	58.3	58.3	41.6	50	45.8	54.2	50	24	18.8	6.3	0	56.3	37.0	42.3	47.1	68.8	74.1	84.6	82.4	37.1	23.8	30	20
3: Stable (unchanged)	10	12.5	12.5	16.7	20	16.7	16.7	25	12	6.3	0	0	9.4	18.5	15.4	11.8	9.4	3.7	3.85	5.9	11.1	4.76	5	0
4: Clinically unimportant deterioration	16.7	4.2	0	0	23.3	20.8	8.3	0	8	0	0	0	21.9	11.1	15.4	23.5	21.9	22.2	11.5	5.9	14.8	19.1	10	26.7
5: Clinically important deterioration	6.7	0	0	0	0	0	0	0	8	0	0	0	0	3.7	7.7	0	0	0	0	0	7.4	4.8	0	6.6

Legend to table 2. Minimum clinically important difference using varying standard errors (MCID-SE); GBS = Guillain-Barré syndrome, CIDP = chronic inflammatory demyelinating polyradiculoneuropathy, SE = standard error, MRC = Medical Research Council, mISS = modified INCAT sensory sumscore, R-ODS = Rasch-built overall disability scale. The white cells reflect clinically important improvement.

Figure 2. Percentage of GBS and CIDP patients reaching clinically important changes applying the concept of minimum clinically meaningful changes using different cut-offs (SE 1x and SE 1.96x) (MCID-SE) in various interval outcome measures



Legend to figure 2. Patients with GBS consistently demonstrated higher scores when compared to patients with CIDP. The R-ODS demonstrated the highest percentages of responders in both GBS and CIDP. MCID-SE = minimum clinically important difference using varying individual standard errors (MCID-SE). GBS = Guillain-Barré syndrome, CIDP = chronic inflammatory demyelinating polyradiculoneuropathy, MRC = Medical Research Council, mISS = modified INCAT sensory sumscore, R-ODS = Rasch-built overall disability scale.

Setting the MCID-SE cut-off at ≥ 1.96 still demonstrated a similar pattern, with a high percentage of clinically important improvement when assessing changes with the R-ODS in GBS (table 2B). A gradual increase of responders was seen in the GBS patients when assessing strength with the Rasch-built MRC sumscore (41.7% reaching the cut-off vs. 83.3% reaching a cut-off of ≥ 1 SE at 12 months of follow-up). The changes in patients with CIDP were lower, with only 6% showing clinically important improvement at 12 months on the mISS in contrast to 25% of the patients with GBS. Between 0-14.3% of patients with CIDP showed clinically important deterioration during follow-up (table 2).

The magnitude of clinically important improvement for all selected measures applied in patients with GBS and CIDP is illustrated in figure 2. As shown, consistently higher values were obtained for GBS patients. Especially the R-ODS reflected clinically important changes in time.

Setting the MCID-SE cut-off at ≥ 1.96 patients with MGUSP and MMN demonstrated the least clinically important improvement (table 3B). Approximately 86% of the patients with MMN demonstrated a clinically unimportant improvement on the R-ODS-MMN. At a MCID-SE cut-off of ≥ 1, clinically important improvement on the R-ODS-MMN was seen in 11.1-37.5% of patients during follow-up (table 3). Using this cut-off of ≥ 1 , half of the patients with MGUSP demonstrated clinically important improvement on the R-ODS at 3 months of whom half still maintained this level of improvement at 12 months of follow-up (table 3A).

Table 3. Categorising the minimum clinically important difference using the variable SE values (MCID-SE) of each longitudinally examined patient with MGUSP and MMN according to the size and direction of changes seen for all outcome measures applied, A: cut-off value at 1x SE and B: cut-off value at 1.96x SE

A: Cut-off value of 1x SE	**MGUSP (n=11) (%) of patients per subgroup**						**MMN (n=10) (%) of patients per subgroup**			
Subgroup classification	Rasch-built MRC sumscore		Rasch-built mISS		R-ODS		Rasch-built modified MRC sumscore		R-ODS-MMN	
Follow-up period (months) ➔	**3**	**12**	**3**	**12**	**3**	**12**	**3**	**12**	**3**	**12**
1: Clinically important improvement	11.1	0	22.2	0	50	25	22.2	37.5	11.1	28.6
2: Clinically unimportant improvement	22.2	0	55.6	66.7	0	25	22.2	0	11.1	57.1
3: Stable (unchanged)	44.4	33.3	11.1	33.3	25	25	33.3	12.5	44.4	14.3
4: Clinically unimportant deterioration	11.1	33.3	11.1	0	25	0	11.1	50	22.2	0
5: Clinically important deterioration	11.1	33.3	0	0	0	25	11.1	0	11.1	0
B: Cut-off value of 1.96x SE	**MGUSP (n=11) (%) of patients per subgroup**						**MMN (n=10) (%) of patients per subgroup**			
Follow-up period (months) ➔	**3**	**12**	**3**	**12**	**3**	**12**	**3**	**12**	**3**	**12**
1: Clinically important improvement	0	0	0	0	37.5	0	0	12.5	11.1	0
2: Clinically unimportant improvement	33.3	0	77.8	66.7	12.5	50	66.7	25	11.1	85.7
3: Stable (unchanged)	44.4	33.3	11.1	33.3	25	25	0	12.5	44.4	14.3
4: Clinically unimportant deterioration	22.2	33.3	11.1	0	25	0	33.3	50	33.3	0
5: Clinically important deterioration	0	33.3	0	0	0	25	0	0	0	0

Legend to table 3. Minimum clinically important difference using varying standard errors (MCID-SE); MGUSP = monoclonal gammopathy of undetermined significance related polyneuropathy, MMN = multifocal motor neuropathy, SE = standard error. MRC = Medical Research Council, mISS = modified INCAT sensory sumscore, R-ODS = Rasch-built overall disability scale. The white cells reflect clinically important improvement.

Discussion

This study illustrates that the Rasch method reveals variable SE values depending on patients' ability levels for the selected impairment and disability interval measures in immune-mediated neuropathies. These variable SE values enable the definition of a responder at the individual person level and may therefore have major implications for future trials in these conditions by defining clinically meaningful change for interval outcome measures.

Although it has been generally accepted to analyse ordinal (Likert) outcome measures as interval measures, the true distance between the different response options remains unknown, thereby making calculations with sumscores less meaningful. In addition, these ordinal outcome measures are used to calculate a group level SE that remains equal for all patients regardless their changing ability level. Modern clinimetric methods like Rasch analysis not only transform ordinal scores into interval scores, but also calculate the precise SE for each person location. The Rasch method demonstrates that the SE is related to the ability level of the patients, thereby varying across the range of the scale.[14, 15]

Furthermore, when establishing a cut-off value for clinically important change, the natural course of the disease needs to be taken into account. For example, the high percentage of GBS and CIDP patients reaching the 1.96 MCID-SE cut-off value for the R-ODS indicate that this outcome measure is capable of capturing the dynamic changes seen during the course and treatment of these illnesses. In addition to general aspects like being valid and reliable, and fulfillment of the Rasch model expectations, the R-ODS has shown to be responsive when using the MCID-SE method for defining a responder in these conditions. Therefore, this outcome measure is proposed as the preferred outcome measure to capture activity limitations and participation restrictions in patients with GBS and CIDP.[25]

In contrast, both MMN and MGUSP in general have a more indolent disease course. Only limited responders could be found at the 1.96xSE cut-off. Therefore, a cut-off value of 1xSE might be more consistent with the clinical course in these conditions. Additionally, sensory change in immune-mediated neuropathies is usually less prominent than motor alteration; as a consequence, the mISS is expected to vary less justifying choosing 1xSE as a cut-off value when using this outcome measure.[32]

As stated previously, the cut-off values of 1xSE and 1.96x SE as minimal clinically important difference (MCID) are defined arbitrarily. Although the change observed with a cut-off of 1xSE could also incorporate measurement error, for patients with an indolent clinical course a cut-off value of 1.96xSE might not capture small but relevant changes as clinically important.

Data of the MRC sumscore and the mISS were analysed using the Rasch method, however, without a complete correction for Rasch requirements such as item bias, local dependency and fit statistics. Even though it was our aim to improve fit to the model, some incomplete corrections of the outcome measures were accepted in order to preserve to the ordinal construction of the scales.

We have shown that the SE is not a fixed value, but varies across the range of an outcome measure enabling the calculation of individual change using the concept of MCID-SE. Furthermore, to determine MCID-SE cut-off values the speed of progression of the disease course and the expected magnitude of relevant change of the quality examined should be considered. Consequently, for physicians a paradigm shift is needed towards the use of outcome measures with a linear construct taking into account the dynamic changing pattern of defining a responder through the concept of MCID using variable SE values. Therefore the MCID-SE method is suggested to be used in future trial in order to establish cut-off values to define a responder in immune-mediated neuropathies.

Based on the findings in this study, the R-ODS questionnaires have demonstrated in addition to fulfilling modern clinimetric requirements, to have relatively high responsiveness patterns in patients with immune-mediated neuropathies compared to other general measures applied. Further exploration of the various responsiveness entities are warranted aiming to establish a core set of outcome measures for standardised use in future trials and follow-up clinical studies in these conditions.

References

1. van Nes, S.I., C.G. Faber, and I.S. Merkies, *Outcome measures in immune-mediated neuropathies: the need to standardize their use and to understand the clinimetric essentials.* J Peripher Nerv Syst, 2008. **13**(2): p. 136-47.
2. van Schaik, I.N., et al., *Intravenous immunoglobulin for multifocal motor neuropathy.* Cochrane Database Syst Rev, 2005(2): p. CD004429.
3. Allen, D., et al., *Treatment for IgG and IgA paraproteinaemic neuropathy.* Cochrane Database Syst Rev, 2007(1): p. CD005376.
4. Hughes, R.A., A.V. Swan, and P.A. van Doorn, *Intravenous immunoglobulin for Guillain-Barre syndrome.* Cochrane Database Syst Rev, 2010(6): p. CD002063.
5. Lunn, M.P. and E. Nobile-Orazio, *Immunotherapy for IgM anti-myelin-associated glycoprotein paraprotein-associated peripheral neuropathies.* Cochrane Database Syst Rev, 2006(2): p. CD002827.
6. Eftimov, F., et al., *Intravenous immunoglobulin for chronic inflammatory demyelinating polyradiculoneuropathy.* Cochrane Database Syst Rev, 2009(1): p. CD001797.
7. Nunnally, J.C., *Psychometric theory*. 1978, New York: McGraw Hill.
8. Merkies, I.S., et al., Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance. J Neurol Neurosurg Psychiatry, 2010. **81**(11): p. 1194-9.
9. Wyrwich, K.W., W.M. Tierney, and F.D. Wolinsky, *Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire.* Qual Life Res, 2002. **11**(1): p. 1-7.
10. Palta, M., et al., *Standard Error of Measurement of 5 Health Utility Indexes across the Range of Health for Use in Estimating Reliability and Responsiveness.* Med Decis Making, 2010.
11. Lin, K.C., et al., *Assessing the stroke-specific quality of life for outcome measurement in stroke rehabilitation: minimal detectable change and clinically important difference.* Health Qual Life Outcomes, 2011. **9**(1): p. 5.
12. Rasch, G., *Probabilistic models for some intelligence and attainment tests.* 1960, Copenhagen: Danmarks Paedagogiske Institut.
13. Tennant, A. and P.G. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* Arthritis Rheum, 2007. **57**(8): p. 1358-62.
14. Hobart, J. and S. Cano, *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.* Health Technol Assess, 2009. **13**(12): p. iii, ix-x, 1-177.
15. Heesch, K.C., L.C. Masse, and A.L. Dunn, *Using Rasch modeling to re-evaluate three scales related to physical activity: enjoyment, perceived benefits and perceived barriers.* Health Educ Res, 2006. **21 Suppl 1**: p. i58-72.
16. *Research criteria for diagnosis of chronic inflammatory demyelinating polyneuropathy (CIDP). Report from an Ad Hoc Subcommittee of the American Academy of Neurology AIDS Task Force.* Neurology, 1991. **41**(5): p. 617-8.
17. Asbury, A.K. and D.R. Cornblath, *Assessment of current diagnostic criteria for Guillain-Barre syndrome.* Ann Neurol, 1990. **27 Suppl**: p. S21-4.
18. *European Federation of Neurological Societies/Peripheral Nerve Society Guideline on management of multifocal motor neuropathy. Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society.* J Peripher Nerv Syst, 2006. **11**(1): p. 1-8.
19. Miescher, G.C. and A.J. Steck, *Paraproteinaemic neuropathies.* Baillieres Clin Neurol, 1996. **5**(1): p. 219-32.
20. Merkies, I.S. and G. Lauria, *131st ENMC International workshop: Selection of Outcome Measures for Peripheral Neuropathy Clinical Trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscul Disord, 2006. **16**(2): p. 149-56.
21. Kleyweg, R.P., F.G. van der Meche, and P.I. Schmitz, *Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barre syndrome.* Muscle Nerve, 1991. **14**(11): p. 1103-9.
22. *Medical Research Council. Aids to the investigation of the peripheral nervous system*. 1943, Her Majesty's Stationary Office: London. p. 1-2.

23. Merkies, I.S., et al., *Psychometric evaluation of a new sensory scale in immune-mediated polyneuropathies. Inflammatory Neuropathy Cause and Treatment (INCAT) Group.* Neurology, 2000. **54**(4): p. 943-9.
24. van Nes, S.I., et al., *Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies.* J Neurol Neurosurg Psychiatry, 2008. **79**(7): p. 832-4.
25. van Nes, S.I., et al., *Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies.* Neurology, 2011. **76**(4): p. 337-45.
26. Merkies, I.S.J., *Evaluation of scales and measurement instruments in immune-mediated polyneuropathies.* 2001, Erasmus Medical Centre: Rotterdam.
27. Andrich, D., et al., *Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0)*, ed. R.L.P. Ltd. 2003, Duncraig, Western Australia.
28. Merkies, I.S., et al., *Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance.* J Neurol Neurosurg Psychiatry, 2010. **81**(11): p. 1194-9.
29. Liang, M.H., *Evaluating measurement responsiveness.* J Rheumatol, 1995. **22**(6): p. 1191-2.
30. Beaton, D.E., M. Boers, and G.A. Wells, *Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research.* Curr Opin Rheumatol, 2002. **14**(2): p. 109-14.
31. Whitesell, J., *Inflammatory neuropathies.* Semin Neurol, 2010. **30**(4): p. 356-64.
32. Merkies, I.S., et al., *Connecting impairment, disability, and handicap in immune mediated polyneuropathies.* J Neurol Neurosurg Psychiatry, 2003. **74**(1): p. 99-104.

Chapter 6

General discussion and summary

General discussion and summary

In this thesis, the outline and results of the first part of the *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study are described. The primary aim of the studies described in this thesis was to improve and standardise assessment of patients with immune-mediated neuropathies with the final aim to present a neuropathy-specific core set of high quality outcome measures for future immune-mediated neuropathy studies (aim of the ***P*eri*N*om*S*** study). Starting with the ***P*eri*N*om*S*** study we could not imagine how much we would be influenced by the changing insights in the field of clinimetrics. This thesis reflects these changing insights by transforming selected ordinal outcome measures to interval measures using the Rasch method before performing the actual comparison studies. Furthermore, changing insights on how to define a responder using the concept of minimum clinically important difference (MCID) influenced the performed comparison studies. Not only in immune-mediated neuropathies, but also in other chronic illnesses the use of the Rasch method and the concept of MCID will be relevant improving assessment of patients.

Chapter 1 (general introduction) highlights the various forms of immune-mediated neuropathies, ranging from an acute onset of motor (-sensory) deficit in Guillain-Barré syndrome (GBS), to more chronic forms such as chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), and (far) more slowly progressive neuropathies such as monoclonal gammopathy of undetermined significance related neuropathy (MGUSP), and multifocal motor neuropathy (MMN). Furthermore, an overview of all outcome measures applied in clinical trials involving patients with these disorders is presented, highlighting their strengths and weaknesses. Since many different outcome measures have been used, thus hampering the comparison of the obtained results, a strong plea for standardisation was made: the use of a standardised set of outcome measures in future studies. The clinimetric essentials, from a classic approach to more modern approaches like item response theory (IRT) and Rasch methodology are discussed.

Finally, the *P*eripheral *N*europathy outcome measures *S*tandardisation (***P*eri*N*om*S***) study is presented, a large clinimetric study carried out by 26 centres worldwide. Its aim is to obtain a core set of neuropathy-specific outcome measures, representing the various levels of assessing outcome, to be used in future clinical studies in immune-mediated neuropathies. The studies underlying this thesis contribute to realising this aim by improving and standardising assessment of patients with immune-mediated neuropathies.

Chapter 2 (normative value studies) presents the revised normative values for the 2-point discriminator and the Jamar dynamometer, after having examined large cohorts of healthy controls. Normative values for the 2-point discriminator were needed to improve the assessment of this quality as part of the INCAT sensory sumscore. Until now arbitrarily chosen cut-off values for the 2-point discrimination quality at the index finger have been used. The revised normative values of the Jamar dynamometer were obtained using similar statistical techniques as have been used to calculate normative values for the Vigorimeter, enhancing comparison.[1]

Chapter 3 (Rasch-built outcome measures), the core of this thesis, reflects a new era in constructing and evaluating outcome measures. It starts with an introduction aiming to explain and simplify the method of Rasch analysis. An example of how to use Rasch to improve traditional ordinal scales by translating them to interval scales is given by the revision of the fatigue severity scale (FSS). Furthermore, the construction of two neuropathy-specific activity and participation measures is described (R-ODS questionnaires), demonstrating the possibilities of using Rasch to create new measures with a linear construct, enabling the use of sumscores. A specific R-ODS was constructed for patients with GBS, CIDP, and MGUSP, fulfilling all Rasch model expectations. Since patients with MMN may substantially differ in their clinical features and as a result differ in activity and participation restrictions, a MMN-specific R-ODS was constructed, also fulfilling all Rasch model expectations. Compared to the currently most often used ordinal scale in immune-mediated neuropathy studies to reflect disability, the overall disability sumscore (ODSS), the R-ODS for GBS, CIDP and MGUSP represents a wider range of item difficulties thereby better targeting patients with different ability levels. For the R-ODS-MMN already some responsiveness analyses are presented (based upon a limited dataset) comparing traditional techniques such as effect size with more dynamic responsiveness analyses using Rasch-generated individual standard errors depending on the ability level of a patient.

Chapter 4 (comparative validity, reliability, and responsiveness) presents the comparison between the Jamar dynamometer and the Vigorimeter for the assessment of grip strength in patients with GBS, CIDP, MGUSP and MMN. Based on the comparative validity and reliability analyses of these two linear measures, no difference was seen between both tools. Their responsiveness needs further evaluation. The preliminary results described in this thesis demonstrated a questionable low ability to detect clinically relevant change

for both tools. However, these results were based on a rather low amount of patients with incomplete follow-up data. These data will be reanalysed when the entire group of longitudinally studied patients included in the ***P*eri*N*om*S*** study has finalised the follow-up phase. Most patients preferred the use of the Vigorimeter over the Jamar dynamometer for various reasons such as 'being less heavy' or 'less painful to squeeze a bulb'. Therefore, if both devices are equally valid, reliable and responsive, the use of the Vigorimeter will be suggested in future clinical studies.

Also, a comparison was made between the sensory subset of the neuropathy impairment score (NISs) and the modified INCAT sensory sumscore (mISS). The ISS was modified by incorporating joint position and touch sense in a standardised manner and improving the 2-point discrimination assessment by using our revised age matched normative values. Both ordinal measures were successfully translated to interval measures fulfilling Rasch requirements in general, and still maintaining their original structure. Despite the apparent similarity, the results of the clinimetric comparative studies are in favour of the mISS, because this scale demonstrated a larger targeting range, less floor effect, and better responsiveness scores. Nevertheless, these analyses will be repeated after closing the longitudinal part of the ***P*eri*N*om*S*** study. Presuming no relevant change when including more data, the use of the mISS is suggested as part of the future standardised core set of outcome measures.

Chapter 5 (minimum clinically important difference (MCID)) introduces the concept of MCID, a concept defining the minimum change score necessary to reflect clinically relevant change. Defining MCID may be complex. There are many different methods to calculate MCID, such as anchor-based and distribution-based methods, however, there is no consensus on the best calculation method. Furthermore, different interested parties (e.g., patients, clinicians, the community) will define MCID in a different manner.

Using data from the largest trial ever performed in patients with CIDP, the ICE study, the efficacy of intravenous immunoglobulin over placebo was reconfirmed, shifting hereby from statistical significant to clinically relevant changes. Since no consensus on the use of MCID methods exists, a suggestion is given to define a responder as a patient that fulfils at least one anchor-based and one distribution-based MCID method. The presented MCID cut-offs derived from the ICE study suggest a static MCID over the whole range of an outcome measure. However, 'defining a responder' seems to have a dynamic changing pattern. As has been reported by others, we noticed that the magnitude of Rasch-generated individual standard errors depend on the ability of a patient reflected by their location on the linear metric used, increasing at the floor and ceiling parts and decreasing at the middle part.[2] Thus, the corresponding MCID cut-off calculated using these individual variable standard errors, the MCID-SE, is also changing per patient. This approach was subsequently applied in the currently available interval measures for patients with immune-mediated neuropathies. Comparative responsiveness studies of these measures were performed by determining the percentage of serially examined patients reaching a pre-defined MCID-SE cut-off. In patients with GBS and CIDP the R-ODS turned out to be more responsive than the Rasch-built MRC sumscore and the Rasch-built mISS.

Main findings described in this thesis

- The fatigue severity scale (FSS), the sensory subset of the neuropathy impairment score (NISs) and the modified INCAT sensory sumscore (mISS), all existing ordinal outcome measures used in patients with immune-mediated neuropathies, were successfully transformed to measures with a linear construct using the Rasch method.
- Two new interval measures capturing activity and participation limitations in immune-mediated neuropathy patients were created using the Rasch method, the Rasch-built overall disability scale (R-ODS) for patients with GBS, CIDP and MGUSP and the R-ODS specifically for patients with MMN (R-ODS-MMN).
- The dynamics of being a responder were demonstrated for Rasch-built outcome measures using individual standard errors provided by Rasch.
- The possibilities of using the concept of minimum clinically important difference (MCID) in defining a responder based on a predefined clinically relevant change instead of a statistical significant change were illustrated with data from the ICE study. The efficacy of intravenous immunoglobulin over placebo in CIDP was reconfirmed using MCID.
- The Jamar dynamometer and the Vigorimeter are both valid and reliable tools. Since their responsiveness seems to be similar as well, the use of the Vigorimeter as the standard tool to assess grip strength in patients with immune-mediated neuropathies is suggested based on patients' preference for the Vigorimeter.
- Clinimetric comparative studies of the sensory subset of the neuropathy impairment score (NISs) and the modified INCAT sensory sumscore (mISS) are in favour of the mISS, since this scale demonstrated a larger targeting range, less floor effect, and better responsiveness scores. However, the comparative responsiveness studies were based on a rather low amount of patients with incomplete follow-up data, thus needs reconfirmation before selecting the mISS for the standardised minimum core set of outcome measures to be used in future neuropathy studies.

- Normative values were provided for the two-point discriminator improving and standardising the mISS. Revised normative values for the Jamar dynamometer were provided to enhance its applicability in clinical practice.

Clinimetric essentials

Outcome measures should be simple, communicable, valid, reliable and responsive prior to their use.[3-5] Currently, clinimetric properties of outcome measures are often examined during clinical trials. However, whether their strict in- and exclusion criteria reflect a representative study population to evaluate these properties is doubtful. To overcome this limitation the ***P*eri*N*om*S*** study includes patients independent from their level of functioning. Furthermore, only simple outcome measures were selected for the comparative ***P*eri*N*om*S*** studies thereby aiming to create a core set of high quality outcome measures not only suitable for future clinical trials but also for daily clinical practice.

Linearity

Ordinal scales based on the classical test theory (CTT) do have disadvantages. Generally, patients are requested to complete all items of CTT-based scales, even though some may be irrelevant for their level of ability. Furthermore, a sum of item scores is often calculated assuming equal relevance and hence weighting of each item which is highly unlikely.[6, 7] The Rasch method overcomes these limitations and enables the construction of interval outcome measures showing the true distance between patients and items on a linear ruler. When all data fulfil the Rasch model expectations meaningful sumscores and change scores can be obtained suitable for conventional statistics.

The calibration of items (item locations on the linear ruler representing their item difficulty) is based on the characteristics of the study population. Some items might be easy to perform for patient with disease A (e.g., MMN), but difficult for patients with disease B (e.g., GBS). Therefore, the construction of disease-specific interval outcome measures should be considered.

Future perspectives

The studies and new insights described in this thesis contribute to improving and standardising assessment of patients with immune-mediated neuropathies. However, there are numerous aspects that still need to be accomplished and these are addressed in the following:

- ***PeriNomS* study:** the longitudinal part is still on-going. Data collection has to be finished before performing the final comparative responsiveness studies. Results of comparative validity and reliability studies of the cross-sectional part will be combined with these responsiveness analyses to present a neuropathy-specific minimum core set of high quality outcome measures for future studies and clinical practice (final aim of the ***P*eri*N*om*S*** study). The comparative responsiveness analyses of the Jamar dynamometer vs. the Vigorimeter and the NISs vs. the mISS will be recalculated.

- **Medical Research Council (MRC) grading system versus NIS motor subset:** as demonstrated for the modified ISS versus NIS sensory subset in chapter 4.2, an equivalent comparison should be part of future clinimetric studies between the MRC grading versus the NIS motor subset using Rasch analyses.

- **MRC grading system:** the MRC grading system, with its 6 scoring options from paralyses (0) to normal strength (5), has served for more than 7 decades as an apparently easy and reliable tool to assess strength of the various muscle groups in neuromuscular illnesses, despite being criticised due to the unequal width of its response categories.[8-11] However, previous reports have suggested humans inability to differentiate between more than 3 to 4 response options.[12, 13] Based on these observations, we examined the applicability and discriminative capacity of physicians using the MRC grades in patients with various neuromuscular illnesses with different degrees of muscle weakness. Discriminative capacity can be visualised through Rasch analyses by examining the ordering of thresholds. Many disordered thresholds were found, therefore a modification of the MRC grading system was suggested.[14]

- **EMG:** its value as part of international guidelines on the diagnosis of various immune-mediated neuropathies has been established. However, there is no consensus whether EMG examination is a useful tool to evaluate treatment effect in clinical trials, and if so, which standardised EMG parameters should be assessed during follow-up for each immune-mediated neuropathy.[15] Further research is necessary to evaluate the relation between improving EMG parameters and relevant clinical improvement from patient's perspective.

- **Skin biopsy:** Skin biopsy as a new diagnostic tool is gaining recognition as a possible outcome measure in patients with peripheral neuropathies.[16] Through skin biopsy, the intraepidermal nerve fibre density (IENFD) can be determined and values can be compared with reported age and gender matched normative cut-off scores.[17] In a recent paper, morphometric studies have demonstrated an expansion of the use of skin biopsy as a possible marker for large nerve fibre neuropathies.[18] Small nerve fibres related dysfunction has been demonstrated in patients with peripheral neuropathies including those with GBS and CIDP. [19-22] Pan and associates were the first that reported on the use of skin biopsy counting the intraepidermal nerve fibre density in patients with GBS and stated that this

illness should be seen as a more global neuropathy, with the involvement of large and small nerve fibres. This has also been investigated in a Dutch prospective clinical study in GBS.[22] However, additional studies are needed to determine the clinical applicability of skin biopsy parameters as possible outcome measures in future clinical trials and to evaluate their prognostic value.

- **MCID:** the concept of minimum clinically important difference (MCID) should have a more central role in defining future goals in trials and follow-up studies in patients with immune-mediated neuropathies, as well as in neurology in general. MCID is a rather complex concept with many ways to define and to calculate its content.[23] Different cut-offs might be necessary for the various neuropathies based on different courses of disease (rapidly progressive versus rather indolent), based on different baseline characteristics or based on the direction of change (improvement versus deterioration). Experts in the field of immune-mediated neuropathies, patient representatives as well as statisticians/clinimetricians should join forces in order to establish the minimum requirements to define clinically meaningful changes for every selected scale of the final core set of outcome measures for each immune-mediated neuropathy separately.

To conclude, the studies described in this thesis can be considered as the basis and the first part of the ***P*eri*N*om*S*** study. When also the longitudinal part of the ***P*eri*N*om*S*** study is finished and analysed, the results will be presented and discussed in an additional workshop in 2012/2013. This will be the final step of the ***P*eri*N*om*S*** study to reach consensus and to present a final minimum core set of high quality outcome measures to be used in future trials and follow-up studies in patients with GBS, CIDP, MGUSP and MMN. The European Neuromuscular Centre (ENMC) could serve as a platform for such a meeting.

References

1. Merkies, I.S., et al., *Assessing grip strength in healthy individuals and patients with immune-mediated polyneuropathies.* Muscle Nerve, 2000. **23**(9): p. 1393-401.
2. Hobart, J. and S. Cano, *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.* Health Technol Assess, 2009. **13**(12): p. iii, ix-x, 1-177.
3. Feinstein, A.R., *Clinimetrics*. 1987, New Haven and London: Yale University Press.
4. Hobart, J.C., D.L. Lamping, and A.J. Thompson, *Evaluating neurological outcome measures: the bare essentials.* J Neurol Neurosurg Psychiatry, 1996. **60**(2): p. 127-30.
5. Nunnally, J.C., *Psychometric theory*. 1978, New York: McGraw Hill.
6. DeVellis, R.F., *Classical test theory.* Med Care, 2006. **44**(11 Suppl 3): p. S50-9.
7. Stucki, G., et al., *Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts.* J Clin Epidemiol, 1996. **49**(7): p. 711-7.
8. Medical Research Council, *Aids to the investigation of the peripheral nervous system*. 1943, London: Her Majesty's Stationanery Office.
9. MacAvoy, M.C. and D.P. Green, *Critical reappraisal of Medical Research Council muscle testing for elbow flexion.* J Hand Surg Am, 2007. **32**(2): p. 149-53.
10. Cuthbert, S.C. and G.J. Goodheart, Jr., *On the reliability and validity of manual muscle testing: a literature review.* Chiropr Osteopat, 2007. **15**: p. 4.
11. Merlini, L., *Measuring muscle strength in clinical trials.* Lancet Neurol, 2010. **9**(12): p. 1146; author reply 1146-7.
12. Andrich, D., *Category ordering and their utility.* Rasch Meas Transact 1996. **9**: p. 464-465.
13. Penta, M., et al., *The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment.* Stroke, 2001. **32**(7): p. 1627-34.
14. Vanhoutte, E.K., et al., *Modifying the Medical Research Council grading system through Rasch analyses.* Brain, in press.
15. Fuglsang-Frederiksen, A. and K. Pugdahl, *Current status on electrodiagnostic standards and guidelines in neuromuscular disorders.* Clin Neurophysiol, 2011. **122**(3): p. 440-55.
16. Lauria, G. and R. Lombardi, *Skin biopsy: a new tool for diagnosing peripheral neuropathy.* BMJ, 2007. **334**(7604): p. 1159-62.
17. Bakkers, M., et al., *Intraepidermal nerve fiber density and its application in sarcoidosis.* Neurology, 2009. **73**(14): p. 1142-8.
18. Lauria, G., et al., *Morphometry of dermal nerve fibers in human skin.* Neurology, 2011. **77**(3): p. 242-9.
19. Pan, C.L., et al., *Cutaneous innervation in Guillain-Barre syndrome: pathology and clinical correlations.* Brain, 2003. **126**(Pt 2): p. 386-97.
20. Chiang, M.C., et al., *Cutaneous innervation in chronic inflammatory demyelinating polyneuropathy.* Neurology, 2002. **59**(7): p. 1094-8.
21. McLeod, J.G., *Autonomic dysfunction in peripheral nerve disease.* J Clin Neurophysiol, 1993. **10**(1): p. 51-60.
22. Ruts, L., *Pain, autonomic dysfunction and course of disease in GBS*. 2010, Erasmus MC: Rotterdam.
23. Beaton, D.E., M. Boers, and G.A. Wells, *Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research.* Curr Opin Rheumatol, 2002. **14**(2): p. 109-14.

Chapter 7

Nederlandse samenvatting

Samenvatting

Hoofdstuk 1 (algemene introductie) belicht de verschillende uitingsvormen van immuun-gemedieerde neuropathieën, variërend van acuut beginnend krachtsverlies (veelal gecombineerd met gevoelsstoornissen) bij het Guillain-Barré syndroom (GBS) tot meer chronische varianten zoals chronische inflammatoire demyeliniserende polyradiculopathie (CIDP), en (zeer) langzaam progressieve neuropathieën zoals polyneuropathie geassocieerd met een monoklonale gammopathie (MGUSP) en multifocale motore neuropathie (MMN). Vervolgens wordt een overzicht gegeven van alle testen en vragenlijsten (uitkomstmaten) die gebruikt zijn in wetenschappelijk onderzoek bij deze patiëntengroepen, inclusief een sterkte-zwakte analyse. Omdat zoveel verschillende uitkomstmaten worden gebruikt, is de vergelijking van onderzoeksresultaten vaak niet goed mogelijk. Daarom wordt gepleit voor standaardisatie: het gebruik van een gestandaardiseerde set van uitkomstmaten in toekomstige onderzoeken. De kwaliteitsvoorwaarden waar deze uitkomstmaten aan moeten voldoen worden toegelicht, waarbij zowel de traditionele begrippen binnen de klinimetrie als ook de plaats voor moderne technieken zoals de item respons theorie (IRT) en de Rasch methode worden besproken.

Uiteindelijk wordt de *P*erifere *N*europathie uitkomstmaten *S*tandaardisatie (***P*eri*N*om*S***) studie gepresenteerd, een grote klinimetrische studie waaraan wereldwijd 26 centra deelnemen. Doel is het vaststellen van een basisset neuropathie-specifieke uitkomstmaten, welke de verschillende niveaus van meten weergeven, en die geschikt is voor de dagelijkse praktijk en voor toekomstig wetenschappelijk onderzoek bij patiënten met een immuun-gemedieerde neuropathie. De studies in dit proefschrift dragen bij aan het realiseren van dit doel door het verbeteren en standaardiseren van het onderzoek van patiënten met een immuun-gemedieerde neuropathie.

Hoofdstuk 2 (normaalwaarden studies) geeft de gereviseerde normaalwaarden voor de 2-puntsdiscriminator en de Jamar dynamometer, verkregen door het onderzoeken van grote groepen gezonde vrijwilligers. Normaalwaarden voor de 2-puntsdiscriminator waren onder meer nodig om de INCAT sensory sumscore te verbeteren. Tot nu toe werden arbitrair gekozen afkappunten gebruikt om het vermogen tot het onderscheiden van 2 punten op de wijsvinger als normaal of abnormaal te classificeren. De gereviseerde normaalwaarden voor de Jamar dynamometer werden verkregen door gebruik te maken van dezelfde statistische technieken als waarmee de normaalwaarden voor de Vigorimeter zijn berekend dat het vergelijken van de twee instrumenten vergemakkelijkt.

Hoofdstuk 3 (uitkomstmaten ontwikkeld met behulp van de Rasch methode), de kern van dit proefschrift, laat nieuwe mogelijkheden zien met betrekking tot het creëren en evalueren van uitkomstmaten. In de introductie wordt de Rasch methode in begrijpelijke taal uitgelegd. De revisie van de fatigue severity scale (FSS) laat zien hoe de Rasch methode gebruikt kan worden om bestaande ordinale schalen om te zetten in intervalschalen. Vervolgens wordt de constructie van twee neuropathie-specifieke vragenlijsten gericht op het meten van beperkingen in dagelijkse en sociale activiteiten beschreven (R-ODS vragenlijsten). Hierbij wordt de Rasch methode ingezet om nieuwe uitkomstmaten met lineaire eigenschappen te creëren. Voordeel hiervan is dat scores opgeteld kunnen worden. Er kon één R-ODS worden gemaakt voor patiënten met GBS, CIDP en MGUSP, welke voldoet aan de eisen van het Rasch model. De klinische verschijnselen van MMN patiënten zijn anders met als gevolg andere beperkingen in dagelijkse en sociale activiteiten, daarom werd een aparte MMN-specifieke R-ODS vragenlijst ontwikkeld welke eveneens voldoet aan alle eisen van het Rasch model. In vergelijking tot de overall disability sumscore (ODSS), de voor wetenschappelijk onderzoek momenteel meest gebruikte ordinale schaal om beperkingen te meten bij patiënten met een immuun-gemedieerde neuropathie, heeft de R-ODS voor patiënten met GBS, CIDP en MGUSP een groter bereik van makkelijke tot moeilijke vragen. Daardoor sluit de R-ODS beter aan bij patiënten met verschillende niveaus van functioneren. In het R-ODS-MMN stuk worden al enige responsiviteitsanalyses getoond, gebaseerd op beperkt beschikbare data. Traditionele technieken als 'effect size' worden vergeleken met moderne, dynamische technieken waarbij met behulp van Rasch verkregen individuele standard errors (meetfouten) worden gebruikt. Deze hangen af van het niveau van functioneren van de patiënt.

Hoofdstuk 4 (vergelijking van validiteit, betrouwbaarheid en responsiviteit) presenteert de vergelijking van de Jamar dynamometer en de Vigorimeter voor het bepalen van de knijpkracht van patiënten met GBS, CIDP, MGUSP en MMN. Op basis van de analyses van de validiteit en de betrouwbaarheid van beide lineaire instrumenten is er geen belangrijk verschil. De responsiviteit van beide instrumenten vergt nader onderzoek. De voorlopige resultaten beschreven in dit proefschrift wijzen erop dat beide instrumenten matig in staat zijn om klinisch relevante veranderingen weer te geven. Dit is echter gebaseerd op data van een relatief klein aantal patiënten waarvan de follow-up gegevens deels onvolledig zijn. Deze analyses zullen dan ook worden herhaald zodra de follow-up fase van de hele groep met longitudinaal gevolgde patiënten geïncludeerd in de ***P*eri*N*om*S*** studie is afgerond. De meeste patiënten gaven de voorkeur aan de Vigorimeter boven de Jamar dynamometer om redenen als 'minder zwaar'

of ‘minder pijnlijk om in ballon te knijpen’. Als beide instrumenten even valide, betrouwbaar en responsief blijken te zijn, dan zal, gezien de voorkeur van de patiënten, het gebruik van de Vigorimeter worden voorgesteld voor toekomstige klinische studies.

Ook worden de resultaten beschreven van de vergelijking tussen het gedeelte van de neuropathy impairment score dat betrekking heeft op gevoelsstoornissen (NISs) en de gemodificeerde INCAT sensory sumscore (mISS). De mISS werd afgeleid van de ISS door toevoeging van de op een gestandaardiseerde wijze onderzochte kwaliteiten positiezin en tastzin en door de score van de kwaliteit 2-puntsdiscriminatie te baseren op onze gereviseerde leeftijdgebonden normaalwaarden. Beide ordinale uitkomstmaten konden met behulp van de Rasch methode worden omgezet naar intervalschalen die aan het Rasch model voldoen, met behoud van hun originele structuur. Ondanks de ogenschijnlijke overeenkomsten, liet vergelijkend onderzoek zien dat de mISS een aantal gunstigere kwaliteiten heeft dan de NISs: de mISS lijkt een betere weerspiegeling te geven van de gevoelsstoornissen zoals aangegeven door deze patiëntengroepen, de ondergrens van de schaal wordt minder snel bereikt en het geeft beter de verandering van de gevoelskwaliteiten in de tijd weer. Ook deze analyses zullen worden herhaald als het longitudinale deel van de ***P*eri*N*om*S*** studie is afgesloten. Indien door het toevoegen van data de resultaten niet noemenswaardig veranderen, dan zal voorgesteld worden de mISS onderdeel te laten uitmaken van het toekomstige basisset van gestandaardiseerde uitkomstmaten.

Hoofdstuk 5 (minimum clinically important difference (MCID)) introduceert het concept MCID, een concept dat betrekking heeft op het definiëren van het verschil in score dat minimaal nodig is om klinisch relevant te zijn. Het kan lastig zijn om de MCID te definiëren. Voor het berekenen van MCID bestaan verschillende methodes, zoals anker-gebaseerde methodes en verdeling-gebaseerde methodes. Er is echter geen consensus over de beste berekeningsmethode. Daarnaast zullen verschillende partijen (bv. patiënten, dokters, de maatschappij) een andere visie hebben op wat een klinisch relevant verschil is.

Door data te gebruiken van de grootste klinische trial ooit uitgevoerd bij patiënten met CIDP, de ICE studie, is opnieuw aangetoond dat intraveneuze behandeling met immunoglobulines superieur is aan placebo behandeling. Dit keer niet alleen op basis van statistisch significante verschillen maar ook door het aantonen van klinisch relevante verschillen met behulp van het concept MCID. Omdat er geen consensus is welke MCID methode het beste kan worden gebruikt wordt voorgesteld om in toekomstige klinische trials patiënten als ‘responsief’ te beschouwen als de opgetreden verandering minstens

voldoet aan 1 anker-gebaseerde en 1 verdeling-gebaseerde afkapwaarde. De gepresenteerde MCID afkapwaarden in de ICE studie suggereren een statische MCID, gelijk over het hele bereik van de uitkomstmaat. Echter het definiëren van een patiënt als wel of niet responsief lijkt een dynamisch patroon te hebben. Zoals al gerapporteerd werd door anderen, merkten ook wij op dat de grootte van de individuele meetfout (standard error = SE), gegenereerd met behulp van Rasch, afhangt van de locatie op de schaal: de SE neemt toe als de schaal zijn ondergrens of bovengrens bereikt, in het middengedeelte is de SE lager. De corresponderende MCID afkapwaarde, berekend met deze variabele individuele SE, de MCID-SE, verandert dus ook per patiënt. Deze aanpak werd vervolgens toegepast op de inmiddels beschikbare intervalschalen voor patiënten met een immuun-gemedieerde neuropathie. De responsiviteit van deze schalen werd vergeleken door het percentage patiënten te bepalen dat per schaal de van tevoren gedefinieerde MCID-SE afkapwaarde bereikte. Bij patiënten met GBS en CIDP bleek de R-ODS vragenlijst responsiever dan de met behulp van Rasch getransformeerde versies van de MRC sumscore en de mISS.

Hoofdstuk 6 (samenvatting en toekomstperspectieven) geeft een samenvatting van de verschillende studies die in dit proefschrift zijn beschreven, de resultaten worden ter discussie gesteld en er worden suggesties gedaan voor toekomstig onderzoek.

Epilogue

List of abbreviations

AAN	American Academy of Neurology
A-CIDP	acute onset CIDP
AIDP	acute inflammatory demyelinating polyradiculoneuropathy
AIDS	acquired immunodeficiency syndrome
ALDS	AMC linear disability score
AMAN	acute motor axonal neuropathy
AMC	Academic Medical Centre Amsterdam
ANOVA	analysis of variance
CASS	composite autonomic scoring scale
CIDP	chronic inflammatory demyelinating poly(radiculo)neuropathy
ClinJSc	clinical judgment score
CMAP	compound muscle action potential
COPM	Canadian occupational performance measure
CSF	cerebrospinal fluid
CTT	classical test theory
DF	degrees of freedom
DIF	differential item functioning
DIP	distal interphalangeal
EFNS	European Federation of Neurological Societies
EMG	electromyography
ENMC	European Neuromuscular Centre
ES	effect size
EuroQoL-5D	EuroQoL group-5D scale
f-score	Hughes' functional grading scale
FSS	fatigue severity scale
GBS	Guillain-Barré syndrome
HIV	human immunodeficiency virus
HMSN	hereditary motor and sensory neuropathy
ICC	intraclass correlation coefficient
ICC	item characteristic curve
ICE trial	IGIV-C CIDP efficacy trial
ICF	International Classification of Functioning, Disability and Health
ICIDH	International Classification of Impairments, Disabilities, and Handicaps
IENF	intraepidermal nerve fibre
IENFD	IENF density
IGIV-C	immune globulin intravenous, 10% caprylate/chromatography purified
INC	Inflammatory Neuropathy Consortium
INCAT	Inflammatory Neuropathy Cause and Treatment
IRT	item response theory
ISS	INCAT sensory sumscore
IVIg	intravenous immunoglobulins
kPa	kilopascal
MADSAM	multifocal acquired demyelinating sensory and motor neuropathy
MAG	myelin-associated glycoprotein
MCID	minimum clinically important difference
MCID-SE	MCID using variable individual standard errors

MCS	mental component score
mD-COMPASS	modified Dutch composite autonomic symptom scale
MGUS	monoclonal gammopathy of undetermined significance
MGUSP	MGUS-related polyneuropathy
mISS	modified ISS
MMN	multifocal motor neuropathy
MRC	Medical Research Council
MS	multiple sclerosis
MTX	methotrexate
NHP	Nottingham health profile
NIS	neuropathy impairment score
NISs	neuropathy impairment score - sensory subset
ODSS	overall disability sumscore
ONLS	overall neuropathy limitations scale
PCS	physical component score
PE	plasma exchange
*P*eri*N*om*S*	*P*eripheral *N*europathy outcome measures *S*tandardisation
PI-NRS	pain intensity - numerical scale
PN	peripheral neuropathies
PNS	Peripheral Nerve Society
PPCM	personal patient-centred measures
PSI	person separation index
QoI	quality of life
RHS	Rotterdam handicap scale
RMC trial	randomised methotrexate CIDP trial
RMI	Rivermead mobility index
R-mISS	Rasch-built mISS
R-NISs	Rasch-built NISs
R-ODS	Rasch-built overall disability scale
R-ODS-MMN	Rasch-built overall disability scale for patients with MMN
RUMM	Rasch Unidimensional Measurement Models
SAIDP	subacute inflammatory demyelinating polyradiculoneuropathy
SD	standard deviation
SDD	smallest detectable difference
SE	standard error
SEM	standard error of measurement
SF-36	short form 36-item health survey
SFFS	short-form fatigue scale
SIP	sickness impact profile
SRM	standardised response mean
VAS	visual analogue scale
VSN	Vereniging Spierziekten Nederland
WEST	Weinstein enhanced sensory test
WHO	World Health Organisation
WHO-QoL bref	short form of the WHO quality of life scale

About the author

Sonja Ingrid van Nes was born on December 11, 1974 in Dordrecht, the Netherlands. She attended secondary school at the 'Christelijk Lyceum/ Thuredrecht college' in Dordrecht, where she graduated in 1994. The same year she started Medical School at Leiden University. During her study she joined a research project reviewing the diagnostic criteria used in studies of reflex sympathetic dystrophy at the department of Neurology of Leiden UMC (research coordinator: prof. dr. J.J. van Hilten). She obtained her Medical degree in 2001.

For one year she worked as a medical doctor in Neuropsychiatry at the Bavo RNO group in Capelle aan de IJssel, under supervision of L. Zegerius (neurologist). Thereafter she switched to clinical neurology working from 2003 to 2004 at the department of Neurology of the 'Groene Hart Ziekenhuis' in Gouda. After travelling a couple of months in Asia, she restarted working in clinical neurology at the department of Neurology of the 'Spaarne ziekenhuis' in Hoofddorp. From 2006 onwards, she is working as a resident in Neurology at Erasmus MC, Rotterdam (head: prof. dr. P.A.E. Sillevis Smitt). She combined her residency with research underlying this thesis at the Department of Neuro-immunology/ Neuromuscular disorders of Erasmus MC under supervision of prof. dr. P.A. van Doorn (Erasmus MC), dr. I.S.J. Merkies (Spaarne ziekenhuis, Hoofddorp) and dr. C.G. Faber (Maastricht UMC).

She lives in Rotterdam, together with Peter van Poppel and their twin daughters Lieke and Bente (2009).

Dankwoord

Velen hebben de afgelopen jaren een bijdrage geleverd aan de totstandkoming van dit proefschrift. Zonder deze hulp was het niet gelukt. Een aantal mensen wil ik hier graag in het bijzonder noemen.

In de eerste plaats gaat mijn dank uit naar mijn copromotoren, dr. I.S.J. Merkies en dr. C.G. Faber. Beste Ingemar, dit klinimetrisch onderzoek is een voortzetting van jouw eerdere studies. Het was prettig voortborduren op die stevige basis. Dank voor het delen van je kennis en energie waarmee je mij telkens weer richting gaf. Beste Karin, ook jou wil ik hartelijk danken voor je betrokkenheid bij het onderzoek en de altijd praktische adviezen.

Vervolgens gaat mijn dank uit naar prof. dr. P.A. van Doorn, mijn promotor. Beste Pieter, doordat jij het belang inzag van dit klinimetrisch onderzoek, kreeg ik de mogelijkheid om het ten uitvoer te brengen. De soms technische papers en presentaties werden met jouw aanwijzingen altijd een stuk toegankelijker. Deel uitmaken van de GBS onderzoeksgroep staat door jou niet alleen garant voor promoveren, maar ook voor allerlei nevenactiviteiten. Dank!

Zonder de leden van de kleine en grote commissie was deze dag niet mogelijk geweest. Daarom mijn dank aan de kleine commissie: prof. dr. R.A.C. Hughes, prof. dr. P.Y.K. van den Bergh en prof. dr. H.J. Stam. Dear professor Hughes, it is a great honour and privilege to have the opportunity to discuss the contents of this thesis with you today. Furthermore, many thanks for the linguistic and typographical corrections made while reading my complete thesis. Prof. dr. P.Y.K. van den Bergh, beste Peter, hartelijk dank voor het beoordelen van het manuscript, het includeren van patiënten en uw kritieken op verschillende papers. Prof. dr. H.J. Stam, hartelijk bedankt voor de snelle beoordeling van het manuscript tussen de drukke werkzaamheden door.
De leden van de grote commissie, prof. dr. M. de Visser en prof. dr. D.W.J. Dippel, wil ik hartelijk danken voor hun tijd om te opponeren.

Prof. dr. P.A.E. Sillevis Smit, beste Peter, hartelijk dank voor de ruimte die je mij geboden hebt om, als een van de laatsten, mijn onderzoek te combineren met de opleiding (AGIKO constructie).

The members of the ***P*eri*N*om*S*** study group, many thanks for all your efforts for the ***P*eri*N*om*S*** study. In particular, I'd like to thank the following inspiring contributors: prof. dr. D.R. Cornblath for always supporting and promoting the ***P*eri*N*om*S*** study; prof. dr. K.C. Gorson for including so many patients and for your constructive comments on several papers; prof. dr. J-M Léger, prof. E. Nobile-Orazio and dr. G. Lauria, thanks for a warm welcome when delivering tools and binders for the ***P*eri*N*om*S*** study and for being always enthusiastic about ***P*eri*N*om*S***; prof. dr. A.F. Hahn, many thanks for your support, it is a pity that you have retired; dr. M. Lunn, dr. D. Bennett, dr. R. Hadden and prof. dr. M. Reilly, never known that IRB procedures could be that restrictive, your contribution is appreciated even more!

Alle patiënten en controle personen wil ik danken voor hun deelname aan de studies. Ook de gastvrijheid bij huisbezoeken was soms hartverwarmend.

Zonder de hulp van onze Utrechtse collega's, dr. N.C. Notermans, prof. dr. L.H. van den Berg, Elisabeth Cats, Jikke-Mien Niermeijer en Jan Stork was het niet mogelijk geweest om voldoende patiënten met MMN en MGUSP te includeren, dank voor de samenwerking.

Alle mede-onderzoekers van kamer Ee2230/de neuro-immunologie vakgroep wil ik danken voor de prettige werksfeer: Karin Geleijns, Ilse Hoppenbrouwers, Marie-Claire de Wit, Rinze Neuteboom, Krista Kuitwaard, Christa Walgaard, Liselotte Ruts, Judith Drenthen, Ellen Maathuis, Naghmeh Jafari, Nadine van der Beek, Immy Ketelslegers, Juna de Vries, Mark Kuijf, Stephan Wens, Ruth Huizinga, Chris Fokke, Bart Jacobs en Esther Brusse.
Krista, Christa en Judith, jullie wil ik in het bijzonder bedanken voor het waarnemen van ***P*eri*N*om*S*** taken als ik afwezig was, altijd bereid, ook ongevraagd.

Mijn Maastrichtse collega's en vooral Els Vanhoutte, Mayienne Bakkers, Mieke Hermans, Martine Peters, Oliver Harschnitz en Rinske Hamers wil ik danken voor de prettige samenwerking. Els, mijn opvolgster, je bent al helemaal 'gerasched', dank en veel succes met het vervolgonderzoek! Mayienne, we zijn samen begonnen in Rotterdam en door het onderzoek werken we nog steeds een beetje samen. Bedankt voor je hulp. Ik gun je ook een mooie afronding van je onderzoek, je kan het! Martine, zonder jou waren er nog geen normaalwaarden voor de Jamar dynamometer, dank voor je inzet. Oliver en Rinske, jullie hebben vele gezonde vrijwilligers onderzocht met de 2-puntsdiscriminator, bedankt. Mieke, hoewel jouw eigen onderzoekstraject een ander onderwerp heeft heb je toch ***P*eri*N*om*S*** klusjes geklaard, typisch Maastrichts, bedankt.

Alle co-auteurs van de artikelen in dit proefschrift wil ik bedanken voor hun bijdrage. I would like to thank all co-authors of the papers in this thesis for their contribution.

Dr. A. Tennant, thanks for your advice on the Rasch analyses.

Prof. dr. H.A. Drexhage, bedankt voor het verlenen van toegang en ondersteuning op het laboratorium. Harm de Wit, dank je voor de hulp bij het samenstellen van de vloeistoffen nodig voor het bewaren/verwerken van de huidbiopten. Elize Haasdijk, bij onverwachte biopten kon ik altijd bij jou aankloppen voor de benodigdheden, erg fijn.

De medewerkers van de Vereniging Spierziekten Nederland wil ik bedanken voor de hulp bij het verzenden van de R-ODS vragenlijsten naar de leden. Prettig dat de VSN de gelegenheid bood voor een dergelijk onderzoek.
Kris Sieradzan, met een cappuccino was alles op te lossen, niet alleen de teleform problemen, dank!
Odette Wissenburg en Ming Wai Wong, altijd waren er weer nieuwe mappen nodig voor de ***P*eri*N*om*S*** centra, dank voor jullie printwerk.

Arjenne Romeijn, dank voor de administratieve ondersteuning van de ***P*eri*N*om*S*** studie.
Minah Ngoredjo, dank voor de planning van onderzoek vs. kliniektijd, AGIKO constructies komen niet voor niets niet veel meer voor.
Karin ter Meulen, dank voor je regelzaken op de 22e.
Het secretariaat van de polikliniek en de kliniek en de verpleging van de afdeling neurologie wil ik bedanken voor hun ondersteuning op de werkvloer.

De moeder van Christiane, de vrouw van Erik en de moeder van Alexandra, bedankt voor (de hulp bij) het vertalen van vragenlijsten.
Els en Jamie Merkies, bedankt voor de fotoshoots ten behoeve van de research manual.

De neurologen en arts-assistenten/physician assistant in het Albert Schweitzer ziekenhuis wil ik danken voor de prettige werksfeer tijdens de afronding van dit proefschrift.

Gelukkig bestaat er meer in het leven dan dit onderzoek, al lijkt dat soms even niet zo. Mijn familie en vrienden wil ik bedanken voor het begrip als ik weer 'even' geen tijd had en voor de getoonde belangstelling, ook al was de inhoud van mijn onderzoek misschien abracadabra.
Jeanine Lagendijk, lieve Nine, altijd geïnteresseerd en scherp. Ook dit deel ik graag met jou, erg fijn zo'n vriendin als paranimf! Martine Nugteren, lieve Tine, altijd een luisterend oor en bemoedigende woorden, dank voor je vriendschap.

Mijn familie ben ik zeer dankbaar. Pap en mam, altijd begripvol en bereid om zo nodig in te springen. Jullie hebben zelfs vrienden en kennissen gestrikt voor een huidbiopt en/of voor het 2-puntsdiscriminatie onderzoek. Janita, mijn grote zus, met jouw academische inslag en abstractievermogen kan ik me geen betere paranimf wensen. Mirjam en Brendan, mijn zusje en haar man, dank voor jullie interesse.

Mijn schoonouders wil ik ook hartelijk danken voor hun steun. Altijd benieuwd naar hoe het er voor stond en jullie hebben zelfs heel wat uren besteed aan het verwerken van onderzoeksvragenlijsten, hartelijk dank.

Lieve Lieke en Bente, alles is relatief door jullie, heerlijk!

Lieve Peter, eindelijk is het boekje af. Zonder jouw steun en vertrouwen had ik het niet kunnen schrijven. Mijn laatste dankbetuiging is daarom helemaal voor jou: DANK VOOR ALLES!

List of publications

Peters M, **van Nes SI**, Vanhoutte EK, Bakkers M, van Doorn PA, Merkies ISJ, Faber CG. Revised normative values for grip strength with the Jamar dynamometer. *J Peripher Nerv Syst. 2011 Mar;16(1):47-50*

van Nes SI, Vanhoutte EK, van Doorn PA, Hermans M, Bakkers M, Kuitwaard K, Faber CG, Merkies ISJ, on behalf of the PeriNomS study group. Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies. *Neurology. 2011 Jan 25;76(4):337-45.*

Merkies IS, **van Nes SI**, Hanna K, Hughes RA, Deng C. Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance. *J Neurol Neurosurg Psychiatry. 2010 Nov;81(11):1194-9.*

van Nes SI, Vanhoutte EK, Faber CG, Garssen M, van Doorn PA, Merkies ISJ, on behalf of the; PeriNomS study group. Improving fatigue assessment in immune-mediated neuropathies: the modified Rasch-built fatigue severity scale. *J Peripher Nerv Syst. 2009 Dec;14(4):268-78.*

van Nes SI, Faber CG, Hamers RM, Harschnitz O, Bakkers M, Hermans MC, Meijer RJ, van Doorn PA, Merkies ISJ, on behalf of the PeriNomS study group. Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies. *J Neurol Neurosurg Psychiatry. 2008 Jul;79(7):832-4.*

van Nes SI, Faber CG, Merkies ISJ. Outcome measures in immune-mediated neuropathies: the need to standardize their use and to understand the clinimetric essentials. *J Peripher Nerv Syst. 2008 Jun;13(2):136-47. Review.*

van Nes SI, Vanhoutte EK, Bakkers M, Kuitwaard K, Drenthen J, Gorson KC, Cornblath DR, Hahn AF, Cats EA, Niermeijer JMF, Notermans NC, van den Berg LH, van Doorn PA, Faber CG, Merkies ISJ, on behalf of the PeriNomS study group. Jamar dynamometer versus Vigorimeter to assess grip strength in immune-mediated neuropathies. *Manuscript in preparation*

van Nes SI, Vanhoutte EK, Bakkers M, Kuitwaard K, Gorson KC, Cornblath DR, Notermans NC, Lewis RA, Hahn AF, Lauria G, Nobile-Orazio E, Devigili G, Padua L, Léger J-M, van den Bergh PYK, Bennett D, Hadden R, Barreira AA, van Doorn PA, Faber CG, Merkies ISJ, on behalf of the PeriNoms study group. Measuring sensory deficit in immune-mediated neuropathies: comparing the sensory subset of the neuropathy impairment score with the modified INCAT sensory sumscore through Rasch analyses. *Manuscript in preparation*

Vanhoutte EK, **van Nes SI**, Cats EA, van der Pol WL, Gorson KC, Gallia F, Bombelli F, Lewis RA, van Doorn PA, van den Bergh PYK, Léger J-M, Nobile-Orazio E, Cornblath DR, van den Berg LH, Merkies ISJ, Faber CG, on behalf of the PeriNomS study group. Rasch-built Overall Disability Scale for multifocal motor neuropathy (RODS-MMN). *Submitted*

Vanhoutte EK**, van Nes SI**, van Doorn PA, Gorson KC, Cornblath DR, Faber CG, Merkies ISJ, on behalf of the PeriNomS study group. Defining a responder: a dynamic changing pattern in immune-mediated neuropathies through Rasch analyses. *Manuscript in preparation*

Vanhoutte EK, Faber CG, **van Nes SI**, Jacobs BC, van Doorn PA, van Koningsveld R, Cornblath DR, van der Kooi AJ, Cats EA, van den Berg LH, Notermans NC, van der Pol WL, Hermans MEC, van der Beek NAME, Gorson KC, Eurelings M, Engelsman J, Boot H, Meijer RJ, Lauria G, Tennant A, Merkies ISJ, on behalf of the PeriNomS group. Modifying the Medical Research Council grading system through Rasch analyses. *Accepted for publication in Brain*

Bakkers M, Faber CG, Drent M, Hermans MC, **van Nes SI**, Lauria G, De Baets M, Merkies IS. Pain and autonomic dysfunction in patients with sarcoidosis and small fibre neuropathy. *J Neurol. 2010 Dec;257(12):2086-90.*

Bakkers M, Merkies IS, Lauria G, Devigili G, Penza P, Lombardi R, Hermans MC, **van Nes SI**, De Baets M, Faber CG. Intraepidermal nerve fiber density and its application in sarcoidosis. *Neurology. 2009 Oct 6;73(14):1142-8.*

Kuitwaard K, van den Berg LH, Vermeulen M, Brusse E, Cats EA, van der Kooi AJ, Notermans NC, van der Pol WL, van Schaik IN, **van Nes SI**, Hop WC, van Doorn PA. Randomised controlled trial comparing two different intravenous immunoglobulins in chronic inflammatory demyelinating polyradiculoneuropathy. *J Neurol Neurosurg Psychiatry. 2010 Dec;81(12):1374-9.*

van de Beek WJ, Schwartzman RJ, **van Nes SI**, Delhaas EM, van Hilten JJ. Diagnostic criteria used in studies of reflex sympathetic dystrophy. *Neurology. 2002 Feb 26;58(4):522-6. Review.*

PhD portfolio

1. PhD training	Year	Workload (ECTS)
General academic skills		
- Introduction to data-analysis	2007	1
- Classical methods for Data-analysis	2007	5.7
- Repeated measurements	2008	1.4
- Biomedical English Writing	2010	2
In-depth courses (e.g. Research school, Medical Training)		
- Introduction to Rasch analysis, University of Leeds	2007	1
- Klinimetrie: het ontwikkelen en evalueren van meetinstrumenten, VUMC	2007	1.5
International conferences and presentations		
- PNS congress, Utah, USA (training session and poster presentation)	2007	2
- PNS meeting, Paris, France (oral presentation and poster presentation)	2008	2
- PNS meeting, Wurzburg, Germany (2 poster presentations)	2009	0.5
- PNS meeting, Sydney, Australia (2 poster presentations)	2010	1
- PNS meeting, Potomac, USA (2 poster presentations)	2011	1
National conferences and presentations		
- Neuromuscular study group, Vaals (oral presentation)	2007	1
- Maastricht-Aachen meeting, Maastricht (oral presentation)	2007	1
- Department of Neurology, Erasmus MC, Rotterdam (oral presentation)	2007	1
- Department of Neurology, Erasmus MC, Rotterdam (oral presentation)	2010	1
Seminars		
- Department journal club and seminars	2006-2011	2
2. Teaching activities	**Year**	**Workload (ECTS)**
Other		
- Reviewing papers for international journals	2010-2011	0.5
Total		**26.1**


5
4
3
2
1
0
-1
-2
-3
-4