



UvA-DARE (Digital Academic Repository)

Modifying the Medical Research Council grading system through Rasch analyses

Vanhoutte, E.K.; Faber, C.G.; van Nes, S.I.; Jacobs, B.C.; van Doorn, P.A.; van Koningsveld, R.; Cornblath, D.R.; van der Kooi, A.J.; Cats, E.A.; van den Berg, L.H.; Notermans, N.C.; van der Pol, W.L.; Hermans, M.C.E.; van der Beek, N.A.M.E.; Gorson, K.C.; Eurelings, M.; Engelsman, J.; Boot, H.; Meijer, R.J.; Lauria, G.; Tennant, A.; Merkies, I.S.J.

DOI

[10.1093/brain/awr318](https://doi.org/10.1093/brain/awr318)

Publication date

2012

Document Version

Final published version

Published in

Brain

License

CC BY-NC

[Link to publication](#)

Citation for published version (APA):

Vanhoutte, E. K., Faber, C. G., van Nes, S. I., Jacobs, B. C., van Doorn, P. A., van Koningsveld, R., Cornblath, D. R., van der Kooi, A. J., Cats, E. A., van den Berg, L. H., Notermans, N. C., van der Pol, W. L., Hermans, M. C. E., van der Beek, N. A. M. E., Gorson, K. C., Eurelings, M., Engelsman, J., Boot, H., Meijer, R. J., ... Merkies, I. S. J. (2012). Modifying the Medical Research Council grading system through Rasch analyses. *Brain*, 135(5), 1639-1649. <https://doi.org/10.1093/brain/awr318>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

OCCASIONAL PAPER

Modifying the Medical Research Council grading system through Rasch analyses

Els Karla Vanhoutte,¹ Catharina Gerritdina Faber,¹ Sonja Ingrid van Nes,² Bart Casper Jacobs,² Pieter Antoon van Doorn,² Rinske van Koningsveld,³ David Reid Cornblath,⁴ Anneke Jelly van der Kooi,⁵ Elisabeth Aviva Cats,⁶ Leonard Hendrik van den Berg,⁶ Nicolette Claudia Notermans,⁶ Willem Lodewijk van der Pol,⁶ Mieke Catharina Elisabeth Hermans,¹ Nadine Anna Maria Elisabeth van der Beek,² Kenneth Craig Gorson,⁷ Marijke Eurelings,⁸ Jeroen Engelsman,⁸ Hendrik Boot,⁸ Ronaldus Jacobus Meijer,⁸ Giuseppe Lauria,⁹ Alan Tennant¹⁰ and Ingemar Sergio José Merkies^{1,8} on behalf of the PeriNomS Study Group*

1 Department of Neurology, Maastricht University Medical Centre, School of Mental Health and Neuroscience, 6202 AZ, Maastricht, The Netherlands

2 Department of Neurology, Erasmus Medical Centre Rotterdam, Rotterdam, 3000 CA, The Netherlands

3 Department of Neurology, Elkerliek Hospital, Helmond, 5700 AB, The Netherlands

4 Department of Neurology, Johns Hopkins Hospital and School of Medicine, Baltimore, MD 21287-7681, USA

5 Department of Neurology, Academic Medical Centre, Amsterdam, 110 DD, The Netherlands

6 The Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, 3508 GA, The Netherlands

7 Department of Neurology, Caritas St. Elizabeth's Medical Centre, Tufts University School of Medicine, Boston, MA 02135, USA

8 Department of Neurology, Spaarne Hospital, Hoofddorp, 2134 TM, The Netherlands

9 Neuromuscular Diseases Unit, IRCCS Foundation, 'Carlo Besta', Milan 20133, Italy

10 Department of Rehabilitation Medicine, Faculty of Medicine and Health, University of Leeds, Leeds LS1 3EX, UK

*The members of PeriNomS Study Group are provided in the Acknowledgements section.

Correspondence to: Ingemar Sergio José Merkies, MD, PhD,
Spaarne Hospital Hoofddorp,
Spaarnepoort 1, 2134 TM,
Hoofddorp, The Netherlands
E-mail: imerkies@spaarneziekenhuis.nl

The Medical Research Council grading system has served through decades for the evaluation of muscle strength and has been recognized as a cardinal feature of daily neurological, rehabilitation and general medicine examination of patients, despite being respectfully criticized due to the unequal width of its response options. No study has systematically examined, through modern psychometric approach, whether physicians are able to properly use the Medical Research Council grades. The objectives of this study were: (i) to investigate physicians' ability to discriminate among the Medical Research Council categories in patients with different neuromuscular disorders and with various degrees of weakness through thresholds examination using Rasch analysis as a modern psychometric method; (ii) to examine possible factors influencing physicians' ability to apply the Medical Research Council categories through differential item function analyses; and (iii) to examine whether the widely used Medical Research Council 12 muscles sum score in patients with Guillain-Barré syndrome and chronic inflammatory demyelinating polyradiculoneuropathy would meet Rasch model's expectations. A total of 1065 patients were included from nine cohorts with the following diseases: Guillain-Barré syndrome ($n = 480$); myotonic dystrophy type-1 ($n = 169$); chronic inflammatory demyelinating polyradiculoneuropathy ($n = 139$); limb-girdle muscular dystrophy ($n = 105$); multifocal motor neuropathy ($n = 102$); Pompe's disease ($n = 62$) and monoclonal gammopathy of undetermined related polyneuropathy ($n = 8$). Medical Research Council data of 72 muscles were collected. Rasch analyses were performed on Medical Research Council data for each cohort separately and

Received April 24, 2011. Revised September 18, 2011. Accepted September 21, 2011. Advance Access publication December 20, 2011

© The Author (2011). Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

after pooling data at the muscle level to increase category frequencies, and on the Medical Research Council sum score in patients with Guillain–Barré syndrome and chronic inflammatory demyelinating polyradiculoneuropathy. Disordered thresholds were demonstrated in 74–79% of the muscles examined, indicating physicians' inability to discriminate between most Medical Research Council categories. Factors such as physicians' experience or illness type did not influence these findings. Thresholds were restored after rescoring the Medical Research Council grades from six to four options (0, paralysis; 1, severe weakness; 2, slight weakness; 3, normal strength). The Medical Research Council sum score acceptably fulfilled Rasch model expectations after rescoring the response options and creating subsets to resolve local dependency and item bias on diagnosis. In conclusion, a modified, Rasch-built four response category Medical Research Council grading system is proposed, resolving clinicians' inability to differentiate among its original response categories and improving clinical applicability. A modified Medical Research Council sum score at the interval level is presented and is recommended for future studies in Guillain–Barré syndrome and chronic inflammatory demyelinating polyradiculoneuropathy.

Keywords: MRC; manual muscle testing; Rasch; neuromuscular disorders

Abbreviations: CIDP = chronic inflammatory demyelinating polyradiculoneuropathy; MRC = Medical Research Council

Introduction

In 2005, a historical essay tracing the history of scoring and summation of neuromuscular weakness as part of daily neurological practice was published by Dyck *et al.* (2005). Mitchell and Lewis (1886) initiated the practice of alphanumeric scoring of neurological signs in the 19th century. However, it was Lovett, an orthopaedic surgeon, who introduced an ordinal scoring of muscle weakness that formed the basis for the Mayo Clinics and Medical Research Council (MRC) manual muscle testing grading systems, of which the MRC system is most widely used (Medical Research Council, 1943; Dyck *et al.*, 2005). Its worldwide recognition is most probably due to its simplicity, and drawings illustrating how limb muscles should be tested. Through the decades, various versions have been published that aimed to improve the methods for muscle examination. The 2010 edition of *Aids to the Investigation of Peripheral Nerve Injuries. Medical Research Council: Nerve Injuries Research Committee* was recently presented on behalf of the guarantors of *Brain*, embracing a historical review and appreciation for its nurtures through the decades (Compston, 2010). Despite being the most cardinal feature of daily neurological practice, the MRC scale has been respectfully criticized due to the unequal width of its categories, with Grades 1, 2 and 3 being too narrow, and 4 being too broad, often leading to attempts to modify the scale (Brandsma *et al.*, 1995; Dyck *et al.*, 2005; Cuthbert and Goodheart, 2007; MacAvoy and Green, 2007; Merlini, 2010).

One of the most common sources of improper use of any outcome measure concerns the inconsistent use of the response options that correspond to the scales' items (Tennant and Conaghan, 2007). This results in what is known as 'reversed or disordered thresholds'. The term threshold refers to the point between two adjacent response categories where either response is equally probable. In the case of the MRC scale, a threshold would be the point between two adjacent categories, such as between MRC Grades 2 and 3. Disordered thresholds occur when physicians have difficulty consistently discriminating between the MRC grades in patients with various degrees of muscle weakness. Surprisingly, no study has systematically examined the

appropriateness of the MRC scale using modern psychometric techniques.

The objectives of this study were: (i) to examine the applicability and discriminative capacity of physicians using the MRC grades in patients with various neuromuscular illnesses with different degrees of muscle weakness. We questioned whether physicians could demonstrate a fairly uniform MRC grades' ordered thresholds pattern along the Rasch scale continuum, since previous reports suggested human's inability to differentiate between more than four response options (Andrich, 1996; Penta *et al.*, 2001); (ii) to investigate the influence of factors possibly affecting the proper use of the MRC grades in clinical practice (such as physician's clinical experience). For these two objectives, the Rasch method as a modern psychometric vehicle was used, solely focusing on threshold and item bias examinations (Rasch, 1960; Tennant and Conaghan, 2007); and (iii) since Guillain–Barré syndrome and chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) are potentially treatable illnesses and the MRC sum score has been often used as an outcome measure to determine efficacy in these illnesses, we have chosen to examine whether this multi-item scale would fulfil all Rasch model expectations in patients with Guillain–Barré syndrome and CIDP and if not, to propose changes to improve its use (Kleyweg *et al.*, 1991; van der Meche and Schmitz, 1992; Merkies, 2001; van Koningsveld *et al.*, 2004; Hughes *et al.*, 2008)

Patients and methods

Patients

The MRC grades of various muscles were collected from different neuromuscular seminal studies published in the last two decades. Most of these studies have guided the worldwide neurological community in understanding the clinical and therapeutic pattern of these illnesses. A total of 1065 patients (Guillain–Barré syndrome: $n = 480$; myotonic dystrophy type 1: $n = 169$; CIDP: $n = 139$; limb-girdle muscular dystrophy: $n = 105$; multifocal motor neuropathy: $n = 102$; Pompe's disease: $n = 62$; and

Table 1 Basic characteristics of patients with neuromuscular disorders

Study/disorder	Patients examined (n)	Age, mean years (SD), range	Gender		Symptoms duration, mean (SD), range (years)
			Female (%)	Male (%)	
INCAT study	113	54.3 (15.1), 14–84	54 (47.8)	59 (52.2)	6.9 (3.1), 0.5–28
Dutch Guillain–Barré syndrome trial, 1992	147	47.5 (19.2), 5–81	71 (48.3)	76 (51.7)	–
Dutch Guillain–Barré syndrome trial, 2004 + Guillain–Barré syndrome pilot study, 1994	250	50.5 (20.1), 7–89	109 (43.6)	141 (56.4)	–
Myotonic dystrophy type-1	169	43.5 (11.5), 18–69	83 (49.1)	86 (50.9)	5.3 (6.9), 0–34
Multifocal motor neuropathy	102	54.3 (12.1), 26–79	76 (74.5)	26 (25.5)	11.8 (8.2), 0.2–43
Pompe's disease	62	48.1 (11.9), 25.1–71.7	29 (46.8)	33 (53.2)	7.9 (9.3), 0–30.5
Limb-girdle muscular dystrophy	105	37.8 (15.6), 3–70	64 (61.0)	41 (39.0)	21.0 (14.5), 0–58
ICE CIDP	117	51.6 (16.5), 18–83	40 (34.2)	77 (65.8)	5.3 (6.2), 0.2–34.3

In the INCAT studies, a total of 113 patients were examined (Guillain–Barré syndrome, $n = 83$; CIDP, $n = 22$ and monoclonal gammopathy-related polyneuropathy of undetermined significance, $n = 8$).

ICE CIDP = immune globulin intravenous for CIDP; INCAT = inflammatory neuropathy cause and treatment.

monoclonal gammopathy related polyneuropathy of undetermined significance: $n = 8$) were included (Table 1 and Supplementary Table 1) (van der Meché and Schmitz, 1992; van der Kooi *et al.*, 1996; de Die-Smulders *et al.*, 1998; Van den Berg-Vos *et al.*, 2002; van Koningsveld *et al.*, 2004; Hagemans *et al.*, 2005; Van Asseldonk *et al.*, 2005; Hughes *et al.*, 2008; Hermans *et al.*, 2010). The initial MRC data of all patients from the above-mentioned cohorts were selected for the purposes of the current study. All patients met their international criteria for their illness (Asbury and Cornblath, 1990; AANA, 1991; Bushby and Beckmann, 1995; Hirschhorn and Reuser, 2001; EFNS/PNS, 2006; Prior, 2009). The diagnosis 'monoclonal gammopathy related polyneuropathy of undetermined significance' was established after excluding all possible causes for the gammopathy and polyneuropathy (Miescher and Steck, 1996). For all studies, consent was obtained according to the Declaration of Helsinki and approval was obtained by the Ethical Committee of the institution in which the original study was performed.

Assessment scale

The MRC grading system provides the following grades: 0, paralysis; 1, only a trace or flicker of muscle contraction is seen or felt; 2, muscle movement is possible with gravity eliminated; 3, muscle movement is possible against gravity; 4, muscle strength is reduced, but movement against resistance is possible and 5, normal strength.

The MRC grades of the following six muscle pairs comprise the MRC sum score for Guillain–Barré syndrome and CIDP: upper arm abductors, elbow flexors, wrist extensors, hip flexors, knee extensors and foot dorsal flexors (Kleyweg *et al.*, 1991). In the remaining cohorts (monoclonal gammopathy of undetermined significance related polyneuropathy, multifocal motor neuropathy, myotonic dystrophy type-1, Pompe's disease and limb-girdle muscular dystrophy), the muscles groups evaluated represented the clinical picture of each illness (see Supplementary Table 1 for available muscles per cohort).

Rasch analysis

Rationale for using Rasch method

In health care, outcome measures often consist of ordinal multi-item questionnaires, based on the classical test theory (DeVellis, 2006). Concerns have been raised about inappropriate analysis of the generated summed scores that are erroneously assumed to be at the interval level (Wright, 1999; Svensson, 2001; DeVellis, 2006). The ability of a scale to provide fundamental measurements should be established before the more commonly reported psychometric attributes such as being simple, valid, reliable and responsive (Tennant *et al.*, 2004a; Tennant and Conaghan, 2007). Modern scientific methods have been adopted to overcome the shortcomings of traditional measurements. One of the most widely used modern techniques is the Rasch method that transforms ordinal obtained scores into interval-level variables, and whose fit of data satisfies numerous checkpoints as part of model expectations (Rasch, 1960; Tennant *et al.*, 2004a; Tennant and Conaghan, 2007).

In the current study setting, the Rasch model assumes that a patient with less weakness (thus more strength) will have a greater chance of receiving a higher MRC grade by the examining physician. A comprehensive description of the Rasch analysis specifically for neurologists is provided elsewhere (Pallant and Tennant, 2007; Tennant and Conaghan, 2007; van Nes *et al.*, 2011). Briefly, the Rasch model shows what should be expected in response to ordinal items if interval scaling is to be achieved. For this, the following criteria should be fulfilled.

- (i) Thresholds examination: when using items with more than two response categories, as for the MRC grades, proper ordering of the response options should be verified using category probability curves for each muscle group examined, since this will reflect the ability of physicians to use the MRC in a correct way (Shaw *et al.*, 1992). Ordered thresholds are where the transition (threshold) between categories map on to the underlying construct in the expected manner. Thus the transition between categories

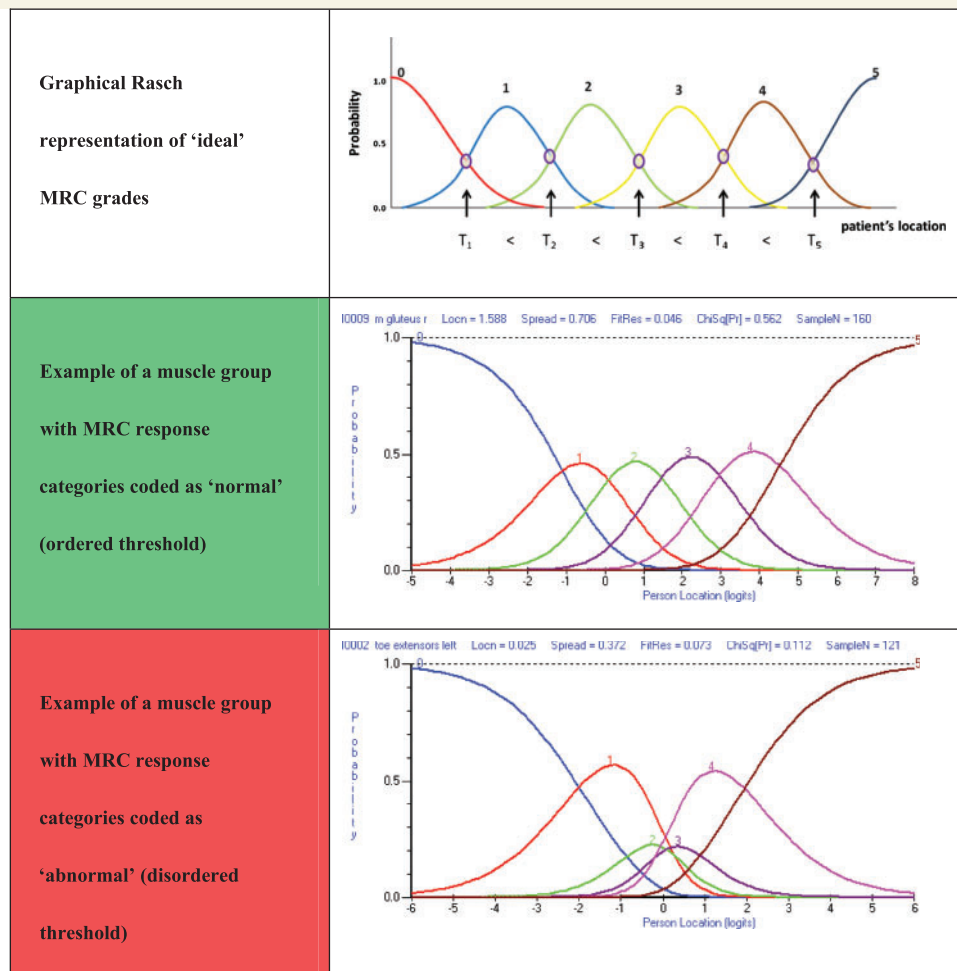


Figure 1 MRC response categories related thresholds explained and coded as 'normal' (green) or 'abnormal' (red). The first row shows the ideal graph representation for proper thresholds for the MRC grades. The first threshold at the intersection between MRC response options 0 and 1 corresponds to a 50% chance of choosing between these two adjacent categories. The thresholds should be ordered to obtain an ideal graph: Threshold 1 < Threshold 2 < Threshold 3 < Threshold 4 < Threshold 5. The second and third row give graphical examples of proper threshold ordering (coded as a green box) and disordered threshold (coded as a red box), respectively. T1–T5 = Thresholds 1–5, respectively.

(e.g. 1–2 and 2–3) reflects increasing levels of muscle strength (Fig. 1, top). Disordered thresholds can occur when physicians use the response options inconsistently, and this inconsistency can be a source of misfit to model expectations. The difficulty discriminating between response options may be a result of too many options, or where the labelling of the options is confusing, both of which may lead to misinterpretation.

- (ii) Fit statistics: fit statistics give an indication of how well the items fit the expected ordering required by the model. This ordering is a probabilistic version of Guttman Scaling (Guttman, 1950). There are two general categories for detecting misfit: overall (summary) misfit, using the entire response matrix, and the individual fit (examining all items and all persons individually). At the summary level the overall mean residual values for both persons and items can be calculated. These values are expressed as a z-score with a mean of 0 and a SD of 1, values of

which indicate perfect fit to model expectations (Tennant *et al.*, 2004b; Pallant and Tennant, 2007). The summary item–trait interaction statistic reflects the fit of the observed data to the model's expectations and is represented by the chi-square. This statistic gives an indication of the invariance of the ordering of items across patients with different levels of muscle strength. A significant chi-square indicates a failure to retain this ordering. Besides the overall fit residuals, individual item–chi square and item and person residuals can be calculated (Tennant *et al.*, 2004b; Pallant and Tennant, 2007; Vandervelde *et al.*, 2007).

- (iii) Item bias: response to an item should not vary between groups (e.g. males versus females), given the same level of the underlying trait (e.g. muscle strength). We assessed item bias (differential item functioning) on the MRC data for various available person factors. A panel (I.S.J.M. and C.G.F.) have studied the range of the factors age, disease duration

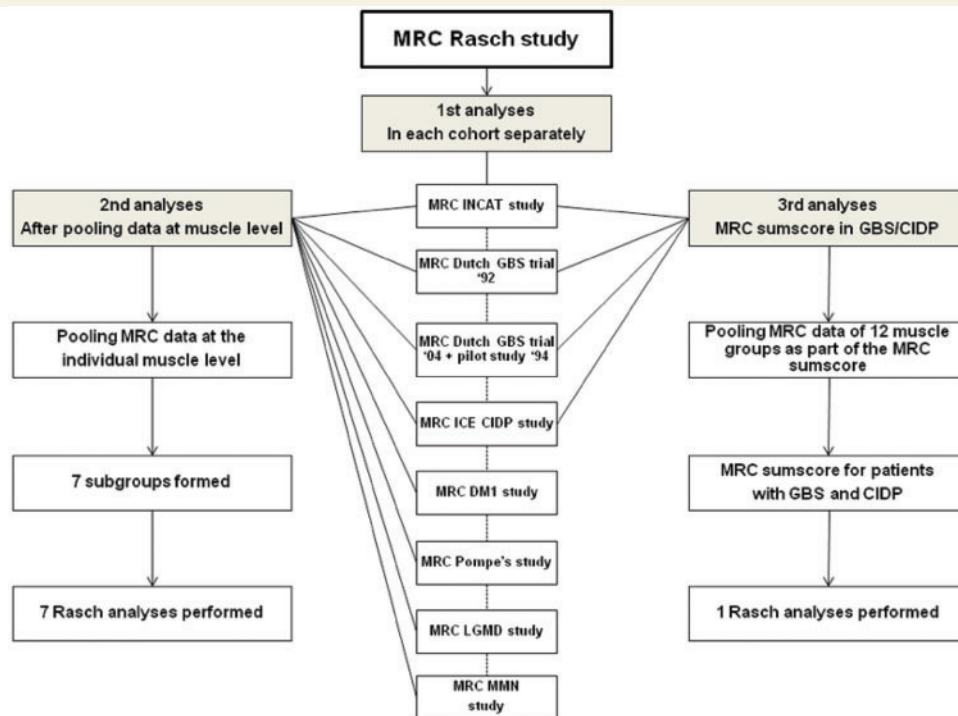


Figure 2 Study algorithm showing a systematic ordering of the analyses performed in the current study. First analyses (Analysis 1): initial MRC Rasch analysis for each individual cohort separately (thus performing a total of eight individual model analyses). Second analyses (Analysis 2): MRC Rasch analyses after pooling data at the muscle level from available cohorts. Third analyses (Analysis 3): MRC sum score Rasch analysis in patients with Guillain–Barré syndrome and CIDP. DM1 = myotonic dystrophy type-1; ICE = immune globulin intravenous for CISP; INCAT = inflammatory neuropathy cause and treatment; LGMD = limb-girdle muscular dystrophy; MMN = multifocal motor neuropathy.

and physician's experience in the available cohorts. Subsequently, these factors were categorized into subgroups for item bias analyses, aiming for an equivalent distribution of participants among the subgroups (25–33% per subgroup).

- (iv) Local dependency: local dependency arises when items are linked such that the response on one item is dependent upon the response to another. Item sets with correlations >0.3 are considered a source of misfit to the model (Tennant and Conaghan, 2007).
- (v) Unidimensionality: the Rasch model assumes unidimensionality and consequently *post hoc* tests are included in the analysis to ensure that this assumption holds. These tests involve a comparison of person estimates (of muscle strength) based upon two sets of items identified from the first principal component analysis of the residuals. The estimates for every individual are compared by a *t*-test, and where $<5\%$ of these comparisons are significantly different, this is taken to support the assumption of unidimensionality (Smith, 2002).

Test procedure

Figure 2 presents a systematic ordering of the analyses performed in the current study. In Analyses 1 and 2 (MRC Rasch analyses for

each cohort separately and MRC Rasch analyses after pooling data) the following were examined:

- Step 1: the presence of ordered thresholds, thus determining whether the MRC grades for each muscle were ordered reflecting physicians' ability to use these grades properly;
- Step 2: in case of disordered thresholds: to seek for the most optimal modified MRC rescored categories that could serve as a unified tool in manual muscle scoring for all muscle groups. In order to rescore the MRC categories, the frequency distribution among the categories and the category probability curves were taken into account;
- Step 3: the presence of possible item bias was examined to determine whether factors such as physician's experience in the neuromuscular field (i.e. would a more experienced physician apply the MRC grades more appropriately than a less experienced physician?) or possible differences between community and university based neurologists might influence the applicability of the MRC grades.

Therefore, in Analyses 1 and 2, the Rasch method was applied only to examine the ability of physicians to use the MRC grading system in a proper way and to determine whether there were factors influencing its use. These analyses were not intended to create a formal Rasch-built MRC sum score for each cohort

individually, since some of the cohort samples were relatively small, hence not fulfilling the basic requirements for proper Rasch modelling (Linacre, 1994).

For Analysis 2, MRC data were pooled at the muscle level from the various available cohorts and resubjected to Rasch analysis, thereby controlling for diagnosis as a possible confounder and strengthening the category frequencies for the various muscles (Linacre, 2002).

In Analysis 3 (MRC sum score Rasch analysis in Guillain–Barré syndrome/CIDP), the MRC 12 muscles sum score was analysed to determine whether Rasch model expectations would be met. The first two steps for Analyses 1 and 2 (see above) were also performed here. Subsequently, since there is no consensus regarding a fixed sequence of steps that must be followed when doing Rasch analyses, our rationale for the following steps were constantly driven by the biggest abnormality seen when studying all subjected data to Rasch, thereby focusing on all aspects that did not meet model expectations (misfit statistics, fit residuals disturbances, under-/over-fitting, local dependency >0.3, and item bias). All steps needed were taken to create a unidimensional scale at the interval level.

Rasch general aspects, person factors and statistics

The MRC data of each muscle group were treated as if it was an 'item' that needed to be completed by the patients with response options from 0 to 5 (in the current study setting: a physician completed the 'item') using the Rasch Unidimensional Measurement Model 2020 software (Andrich *et al.*, 2003).

In Analysis 1 (MRC Rasch in each cohort separately), the following person factors were taken into account (Supplementary Table 2):

- (i) Ages: 1, <40 years; 2, 40–59 years and 3, \geq 60 years;
- (ii) gender: 0, female; 1, male;
- (iii) type of disease: (a) inflammatory neuropathy-cause-and-treatment cohort: 1, Guillain–Barré syndrome; 2, CIDP; 3, gammopathy related polyneuropathy; (b) myotonic dystrophy cohort: 1, mild; 2, adult; 3, child/congenital type; and (c) limb-girdle dystrophy cohort: 1, sarcoglycanopathy; 2, calpainopathy; 3, limb-girdle type 1B, 2B and 2I; 4, unclassified;
- (iv) duration of disease: (a) for all cohorts except limb-girdle patients: 1, <5 years; 2, 5–9 years; 3, 10–19 years; 4, \geq 20 years; and (b) for limb-girdle cohort: 1, <10 years; 2, 10–19 years; 3, 20–29 years; 4, \geq 30 years;
- (v) physician's experience in the neuromuscular field: for the inflammatory–neuropathy cause and treatment studies: 1, <3 years experience; 2, 3–5 years experience; 3, \geq 6 years experience; the latter group constituting senior neuromuscular experts;
- (vi) institution; for the Guillain–Barré syndrome trials: 0, community based; 1, university based hospital; and
- (vii) country; for the Guillain–Barré syndrome cohort 2004: 1, The Netherlands; 2, Germany; 3, Belgium.

For Analyses 2 and 3 (MRC Rasch after pooling data and MRC sum score in Guillain–Barré syndrome/CIDP), the factors studied

included (i) age category: 1, <40 years; 2, 40–59 years; 3, \geq 60 years; (ii) gender: 0, female; 1, male; and (iii) type of disease: depending on the amount of illnesses being pooled together, each illness received a separate code.

For the MRC sum score analysis, the person separation index was also determined, which should be \geq 0.7 for proper group comparison, and a minimum of 0.9 for clinical use (Bland and Altman, 1997). The unrestricted partial credit Rasch model was used. Further analyses were undertaken using Stata 11.0 statistical software for Windows XP.

Results

General aspects

A total of 1065 patients with various neuromuscular disorders were included from nine studies. Table 1 presents the patients' characteristics. MRC data on 72 muscle groups were available (Supplementary Table 1, muscle groups assessed per cohort).

Analysis 1: initial MRC Rasch analyses for each cohort separately

Step 1: thresholds examination

The obtained data (ordered thresholds coded 'green'; disordered coded 'red'; see Fig. 1 explaining these codes) for each muscle group in each cohort were summed, thereby creating a total of 210 muscle groups examined. A total of 165 (78.6%) muscle groups had disordered thresholds versus 45 (21.4%) with ordered thresholds. The disordered thresholds were particularly seen in the mid-response MRC category area (options 2 to 4).

Step 2: rescaling MRC categories

A panel of neuromuscular and Rasch researchers studied the category probability curves and category frequencies of the MRC data for each muscle group. Subsequently, all muscle groups were systematically rescaled in order to obtain the maximum uniform amount of response options, which turned out to be four categories (instead of six). Of the 210 muscle groups rescaled, 182 (86.7%) had ordered thresholds and 28 (13.3%) were still disordered. Sixteen of these disordered muscle groups were distally located (finger spreaders, flexors and extensors, grip strength, wrist extensors and flexors, foot dorsal and plantar flexors). All disordered muscle groups except two were found in the two cohorts with the lowest number of patient's records (multifocal motor neuropathy, $n = 102$ and Pompe's disease, $n = 62$).

Step 3: item bias examination

Eight selected person factors were used to examine possible item bias on the available muscle groups (see Supplementary Table 2 for available factors per cohort). Before rescaling, a total of 806 muscle groups (96.9%) were free of item bias, thus not being influenced by person factors like physicians' experience. Item bias was only found in 26 muscles (3.1%; on person factor gender: 11 muscle groups had uniform differential item

Table 2 Results before and after rescoring the response options from six to four categories with corresponding threshold locations

	Studies (n)	Pooling MRC data per muscle group Patients examined (n)	Before rescoring	After rescoring	Thresholds location		
					Threshold 1	Threshold 2	Threshold 3
Face, neck and trunk muscles			Face, neck and trunk muscles				
Neck flexors	3	336	Normal	Normal	-3.264	1.114	4.919
Neck extensors	3	336	Abnormal	Normal	-4.159	-0.561	1.952
Sternocleidomastoid (right)	2	167	Abnormal	Normal	-7.343	-2.472	0.074
Sternocleidomastoid (left)	2	167	Abnormal	Normal	-7.356	-2.429	0.013
Trapezius (right)	2	167	Abnormal	Normal	0.748	1.642	2.459
Trapezius (left)	2	167	Abnormal	Normal	0.487	1.762	2.452
Pectoralis (right)	2	167	Abnormal	Normal	-1.001	1.644	4.24
Pectoralis (left)	2	167	Abnormal	Normal	-1.059	1.475	4.281
Maseter (right)	1	105	Abnormal	Abnormal			
Maseter (left)	1	105	Abnormal	Abnormal			
Infraspinatus (right)	1	105	Abnormal	Normal	-0.448	2.754	6.063
Infraspinatus (left)	1	105	Abnormal	Normal	-0.169	2.354	6.142
Rhomboid (right)	1	105	Abnormal	Normal	0.001	3.012	7.695
Rhomboid (left)	1	105	Abnormal	Normal	0.028	2.819	7.72
Back extensor muscles	1	105	Abnormal	Normal	1.819	2.858	11.714
Latissimus dorsi (right)	1	105	Abnormal	Normal	0.637	2.691	11.909
Latissimus dorsi (left)	1	105	Abnormal	Normal	0.442	3.143	11.904
Abdominal muscles	1	105	Abnormal	Normal	1.894	2.611	11.872
Upper extremity muscles			Upper extremity muscles				
Shoulder abductors (right)	7	1062	Abnormal	Normal	-3.232	-0.683	2.643
Shoulder abductors (left)	7	1062	Abnormal	Normal	-2.99	-0.727	2.668
Arm exorotation (right)	1	62	Abnormal	Normal	-7.252	-0.432	7.684
Arm exorotation (left)	1	62	Abnormal	Normal	-7.247	-0.442	7.689
Shoulder adductors (right)	1	62	Abnormal	Normal	-6.99	-0.293	7.283
Shoulder adductors (left)	1	62	Abnormal	Normal	-6.984	-0.307	7.291
Elbow flexors (right)	7	1062	Abnormal	Normal	-3.454	-0.441	2.672
Elbow flexors (left)	7	1062	Abnormal	Normal	-3.212	-0.547	2.526
Elbow extensors (right)	4	438	Abnormal	Normal	-4.118	-1.027	2.926
Elbow extensors (left)	4	438	Normal	Normal	-4.038	-1.258	2.796
Wrist extensors (right)	7	1062	Abnormal	Normal	-3.015	0.267	2.66
Wrist extensors (left)	7	1062	Abnormal	Normal	-3.471	0.348	2.685
Wrist flexors (right)	3	269	Normal	Normal	-5.606	-1.149	2.615
Wrist flexors (left)	3	269	Abnormal	Normal	-4.333	0.368	2.159
Brachioradial (right)	1	105	Abnormal	Normal	-3.236	0.691	3.532
Brachioradial (left)	1	105	Abnormal	Normal	-3.51	-0.128	3.189
Pronator (right)	1	105	Abnormal	Normal	-3.595	-0.29	6.767
Pronator (left)	1	105	Abnormal	Normal	-3.521	-0.383	6.771
Supinator (right)	1	105	Abnormal	Normal	-3.905	0.966	6.373
Supinator (left)	1	105	Abnormal	Normal	-3.865	0.707	6.399
Grip (right)	2	167	Abnormal	Normal	-7.509	-4.424	-3.069
Grip (left)	2	167	Abnormal	Normal	-7.284	-4.55	-2.852
Finger extensors (right)	3	269	Abnormal	Normal	-2.376	-1.097	3.475
Finger extensors (left)	3	269	Abnormal	Normal	-2.704	-0.378	3.38
Fingers spreaders (right)	3	269	Abnormal	Normal	-3.185	0.76	5.329
Finger spreaders (left)	3	269	Abnormal	Normal	-3.459	1.21	4.993
Finger flexors (right)	3	376	Normal	Normal	-5.181	-0.236	5.701
Finger flexors (left)	3	376	Normal	Normal	-5.462	0.064	5.114
Thumb abductor (right)	1	102	Abnormal	Normal	0.216	2.122	4.752
Thumb abductor (left)	1	102	Abnormal	Normal	-0.79	2.448	5.209
Thumb adductor (right)	1	102	Normal	Normal	-4.886	0.752	3.575
Thumb adductor (left)	1	102	Abnormal	Normal	-3.942	-0.016	3.38
Opponens pollicis (right)	1	102	Abnormal	Normal	-4.834	0.631	3.836
Opponens pollicis (left)	1	102	Abnormal	Normal	-4.623	0.423	3.89

(continued)

Table 2 Continued

	Studies (n)	Pooling MRC data per muscle group Patients examined (n)	Before rescoring	After rescoring	Thresholds location		
					Threshold 1	Threshold 2	Threshold 3
Lower extremity muscles			Lower extremity muscles				
Gluteus (right)	2	167	Normal	Normal	−2.21	1.781	5.175
Gluteus (left)	2	167	Normal	Normal	−2.2	1.703	5.337
Hip flexors (right)	7	1062	Normal	Normal	−2.394	0.152	3.055
Hip flexors (left)	7	1062	Normal	Normal	−2.295	0.076	3.045
Hip abductors (right)	2	167	Normal	Normal	−2.295	0.681	4.134
Hip abductors (left)	2	167	Normal	Normal	−2.299	0.663	4.214
Hip adductors (right)	2	167	Abnormal	Normal	0.038	2.138	5.434
Hip adductors (left)	2	167	Abnormal	Normal	0.135	2.223	5.418
Knee flexors (right)	4	438	Normal	Normal	−1.333	0.92	4
Knee flexors (left)	4	438	Normal	Normal	−1.3	0.795	4.156
Knee extensors (right)	7	1062	Normal	Normal	−2.721	−0.418	2.064
Knee extensors (left)	7	1062	Normal	Normal	−2.739	−0.381	2.052
Foot dorsal flexors (right)	7	1062	Abnormal	Normal	−1.589	1.463	3.013
Foot dorsal flexors (left)	7	1062	Abnormal	Normal	−1.578	1.408	3.093
Foot plantar flexors (right)	4	438	Normal	Normal	−4.278	−0.166	3.273
Foot plantar flexors (left)	4	438	Abnormal	Normal	−4.218	−0.339	3.209
Toes extensors (right)	2	207	Abnormal	Normal	−2.459	0.21	2.048
Toes extensors (left)	2	207	Abnormal	Normal	−2.271	−0.003	2.474
Toes flexors (right)	1	102	Abnormal	Normal	−4.328	0.921	2.41
Toes flexors (left)	1	102	Abnormal	Normal	−4.194	0.954	2.497

A normal threshold ordering of the MRC grades is coded as 'normal'; abnormal threshold is 'abnormal'. See Fig. 1, for examples, explaining these codes. Threshold location = location of the thresholds of adjacent MRC response options located on the created ruler (and expressed in logits).

functioning, on disease type: eight had uniform, on disease duration: two uniform and one non-uniform, on physician's experience: two uniform, on country: one uniform, and on age: one muscle group had uniform differential item functioning). Differential item functioning findings did not change after rescoring at the individual cohort level.

Analysis 2: MRC Rasch analyses after pooling data

Step 1: thresholds examination

Similar findings were seen in the pooled data analyses. Of the 72 muscles examined, a total of 53 muscle groups (73.6%) had disordered threshold, particularly in the mid-categories (Table 2, 'before rescoring').

Step 2: rescoring MRC categories

Equivalent to the findings of Analysis 1 and based on the location seen of the disordered thresholds (mid-categories 2–4), all muscle groups were systematically rescored to a modified MRC with four categories. Table 2 provides the data for the rescored MRC categories (see last four columns). Ordered thresholds were restored for all muscles except the masseter muscle. A modified version of the MRC grading system was created for clinical use with the following grades: 0, paralysis; 1, severe weakness (defined as >50% loss of strength); 2, slight weakness (<50% loss of strength); and 3, normal strength. A 50% cut-off was based on the following: having four modified response options as having

three thresholds (three theoretical intersections between adjacent response options: Thresholds 1, 2 and 3); half of the distance between Threshold 3 (representing the intersection between modified MRC Grades 2 and 3; location 4.3 logits) and Threshold 1 (intersection between modified grades 0 and 1; location −2.98) for all 72 muscle groups is located at 0.66 logits [−2.98 (location Threshold 1) + 0.5 × 7.28 (0.5 × distances between Threshold 3 and Threshold 1)], which is close to the mean for Threshold 2 (intersection between the modified Grades 1 and 2): 0.46.

Step 3: item bias examination

Differential item functioning was also performed on person factors age, gender and diagnosis (Supplementary Table 3). Item bias was hardly seen on age and gender. On diagnosis, 33 muscle groups (45.8%) demonstrated differential item functioning (Supplementary Table 3).

Analysis 3: MRC sum score Rasch analysis in patients with Guillain–Barré syndrome and chronic inflammatory demyelinating polyradiculoneuropathy

Step 0: general description of patients examined and initial findings

A total of 619 patients from several cohorts [Guillain–Barré syndrome, $n = 480$; CIDP, $n = 139$; $n = 272$ females (43.9%) and $n = 347$ males (56.1%)] were available for these analyses

Table 3 Summary Rasch analyses statistics for the modification of MRC sum score in patients with Guillain–Barré syndrome and CIDP

Analysis	Item fit residuals	Person fit residuals	Item-trait chi-square interaction		PSI	Unidimensionality independent t-test (95% CI)
	Mean (SD)	Mean (SD)	DF	P-value		
Initial	0.147 (4.626)	−0.562 (1.749)	108	<0.00001	0.94	0.20 (0.183–0.218)
Final	0.341 (1.100)	−0.316 (1.094)	55	0.0891	0.91	NA

In the final analysis, item and person fit residuals are acceptable, whereas chi-square is non-significant, indicating invariance across the trait. A person separation index of 0.91 indicates a reliable internal consistency. NA = not available; after performing split analyses, Rasch Unidimensional Measurement Model does not provide the opportunity to perform unidimensionality testing.

DF = degrees of freedom; PSI = person separation index.

(van der Meche and Schmitz, 1992; The Dutch Guillain–Barré syndrome study group, 1994; Merckies, 2001; van Koningsveld *et al.*, 2004; Hughes *et al.*, 2008). The original MRC summed score failed to meet the model expectations. Misfit statistical findings for all three statistical parameters were initially seen (Table 3, ‘initial’ analysis).

Steps 1 and 2: thresholds examination and rescoring

Similar findings were seen here as the above-mentioned analyses. Eight muscle groups had disordered threshold. For uniformity, all 12 muscle groups were rescored to four response options, thereby restoring threshold ordering.

Step 3: local dependency and creating subsets

The following steps were driven by the strongest misfit seen to the Rasch model, which was found to be the strong local dependency findings of equivalent (right and left) muscle pairs (e.g. shoulder abductors right and left side; Spearman’s correlations: $\rho = 0.676–0.831$). Therefore, six subsets of items were created, by combining the corresponding muscle pairs (left and right) with each other, improving the statistical parameters and resolving local dependency.

Step 4: unidimensionality examination

Based on the first principal components analysis, two comparison groups of subsets were formed with three positively loaded (arm muscle subsets) versus three negatively loaded (leg muscle subsets). The independent *t*-tests between these two groups suggested acceptable unidimensionality [*t*-test (95% confidence interval): 0.065 (0.047–0.082)].

Step 5: item bias examination

Uniform differential item functioning was demonstrated on person factor ‘disease type’ for all created muscle subsets, except for the elbow flexors subset. Therefore, each subset of muscle pairs was split in order to obtain specific subsets for the patients with Guillain–Barré syndrome and CIDP, separately. After this, the model was free of any item bias and local dependency. All subsets of items, except the ‘foot dorsal flexors for patients with Guillain–Barré syndrome’, demonstrated fit statistics within required limits. The foot dorsal flexors in Guillain–Barré syndrome had a fit residual of +5.845 ($P = 0.000021$), which disturbed Rasch model fitting (Table 3, final analysis for complete model fit after

removing this item). However, for practical reasons the structure of the MRC sum score (composed by 12 muscles) was maintained, despite having skewed foot dorsal flexors in the Guillain–Barré syndrome subset of item. A high person separation index (0.91) was obtained for the final modified MRC sum score model.

Discussion

Manual muscle testing has been used for more than seven decades for monitoring disease progression and response to therapy in various neuromuscular disorders (van der Meche and Schmitz, 1992; van der Kooi *et al.*, 1996; de Die-Smulders *et al.*, 1998; Merckies, 2001; Van den Berg-Vos *et al.*, 2002; van Koningsveld *et al.*, 2004; Hagemans *et al.*, 2005; Van Asseldonk *et al.*, 2005; Hughes *et al.*, 2008; Hermans *et al.*, 2010) and the MRC grading system has been widely used for this purpose (Dyck *et al.*, 2005; Compston, 2010). This study systematically examined the discriminatory capacity of the MRC grading system in a broad mixture of patients with neuromuscular illnesses, assessing a large number of muscles using the Rasch method. The original six response categories of the MRC grading system failed to differentiate among patients with various degrees of muscle weakness. Three-quarters of all muscles examined demonstrated disordered thresholds, especially in the mid-response categories (options 2–4). The inability of physicians to apply the apparently intuitive and easily applicable MRC grades in a proper way is consistent with reports criticizing the MRC system (Dyck *et al.*, 2005; Schreuders *et al.*, 2006; MacAvoy and Green, 2007; Merlini, 2010). The current paper also shows that the observed disordered thresholds were generally independent of factors such as physicians’ experience, duration of illness or type of practice (university- versus community-based). The original MRC grading system inconsistencies were also ‘cross-validated’ throughout the neuromuscular cohorts, as the findings between the individual disease cohorts were equivalent.

After systematically rescoring all MRC grades to a modified four category response option, the accuracy of the MRC grading system increased by fulfilling ordered thresholds requirements. While this change from six to four response options might intuitively lower the ability to capture functional changes in a patient, from the current evidence, however, keeping the six responses will give a false sense of precision and potentially increase the error in

assessment, which may lead to a false sense of clinically meaningful improvement when it may not exist.

The current paper shows the difficulties with the use of summed scores derived from various muscles tested in patients with Guillain–Barré syndrome and CIDP. However, after Rasch modelling, we were able to present a transformed modified MRC 12 muscle groups summed score for use in future clinical studies in these disorders (Kleyweg *et al.*, 1991). The analyses revealed severe misfit of the foot dorsal flexors. However, since Guillain–Barré syndrome and CIDP are length-dependent neuropathies, we decided to keep this muscle group in the final model. The presented Rasch-built modified interval MRC sum score is considered a substantial improvement compared to the evaluation of muscle strength using ordinal based scores, which in essence are not suitable for performing adequate statistics. The modified interval MRC sum score for patients with CIDP should, however, be applied with some caution, because only 139 patients were assessed, which is lower than the proposed sample size requirements for a stable model (Linacre, 1994). Also, the responsiveness of the Rasch-built modified interval MRC summed score for patients with Guillain–Barré syndrome and CIDP needs to be demonstrated in longitudinal studies, which is currently being investigated (Liang, 1995). However, its personal separation index was high, indicating good ability of the modified scale to differentiate between groups of patients with various degrees for muscle weakness. Finally, since the differential item functioning findings on diagnosis (Supplementary Table 3) demonstrate that neuromuscular illnesses may behave differently, it is conceivable that Rasch-built MRC sum scores are needed for specific illnesses such as multifocal motor neuropathy and other neuromuscular diseases. These efforts should be the focus of future studies.

In conclusion, the original MRC manual muscle testing grading system failed to meet the Rasch model expectations in various neuromuscular disorders, despite being the standard metric in neurology worldwide. Modification of this grading system to four response categories (0, paralysis; 1, severe weakness; 2, slight weakness; and 3, normal strength) may significantly enhance the ability of clinicians to differentiate degrees of weakness with greater precision and accuracy. Based on this, we have developed a Rasch-built interval MRC summed score for use in future clinical studies evaluating patients with Guillain–Barré syndrome and CIDP. Future studies are warranted to improve the solidness of our neurological assessments.

Supplementary Material

Supplementary material is available at *Brain* online.

Acknowledgements

We thank Professor S. Waxman from the Yale University, USA who helped us to increase the transparency and reading of the manuscript. The members of PeriNomS Study Group are as follows: A.A. Barreira, Brazil; D. Bennett, UK; P.Y.K. van den Bergh, Belgium; V. Bril, Canada; G. Devigili, Italy; R.D. Hadden, UK;

A.F. Hahn, Canada; H.-P. Hartung, Germany; R.A.C. Hughes, UK; I. Illa, Spain; H. Katzberg, Canada; A.J. van der Kooij, The Netherlands; J.-M. Léger, France; R.A. Lewis, USA; M.P.T. Lunn, UK; O.J.M. Nascimento, Brazil; E. Nobile-Orazio, Italy; L. Padua, Italy; J. Pouget, France; M.M. Reilly, UK; I. van Schaik, The Netherlands; B. Smith, USA; M. de Visser, The Netherlands; D. Walk, USA

References

- AANA. 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: 617–8.
- Andrich D. Category ordering and their utility. *Rasch Meas Transact* 1996; 9: 464–65.
- Andrich D, Luo G, Sheridan BE, Lyne A, Sheridan B. Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0). Duncraig. Western Australia: RUMM Laboratory; 2003.
- Asbury AK, Cornblath DR. Assessment of current diagnostic criteria for Guillain–Barré syndrome. *Ann Neurol* 1990; 27 (Suppl): S21–4.
- Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997; 314: 572.
- Brandsma JW, Schreuders TA, Birke JA, Piefer A, Oostendorp R. Manual muscle strength testing: intraobserver and interobserver reliabilities for the intrinsic muscles of the hand. *J Hand Ther* 1995; 8: 185–90.
- Bushby KM, Beckmann JS. The limb-girdle muscular dystrophies—proposal for a new nomenclature. *Neuromuscul Disord* 1995; 5: 337–43.
- Compston A. Aids to the investigation of peripheral nerve injuries. Medical Research Council: Nerve Injuries Research Committee. His Majesty's Stationery Office: 1942; pp. 48 (iii) and 74 figures and 7 diagrams; with aids to the examination of the peripheral nervous system. By Michael O'Brien for the Guarantors of Brain. Saunders Elsevier; 2010. pp. [8] 64 and 94 Figures. *Brain* 2010; 133: 2838–44.
- Cuthbert SC, Goodheart GJ Jr. On the reliability and validity of manual muscle testing: a literature review. *Chiropr Osteopat* 2007; 15: 4.
- de Die-Smulders CE, Howeler CJ, Thijs C, Mirandolle JF, Anten HB, Smeets HJ, *et al.* Age and causes of death in adult-onset myotonic dystrophy. *Brain* 1998; 121 (Pt 8): 1557–63.
- DeVellis RF. Classical test theory. *Med Care* 2006; 44 (11 Suppl 3): S50–9.
- Dyck PJ, Boes CJ, Mulder D, Millikan C, Windebank AJ, Dyck PJ, *et al.* History of standard scoring, notation, and summation of neuromuscular signs. A current survey and recommendation. *J Peripher Nerv Syst* 2005; 10: 158–73.
- EFNS/PNSEuropean 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: 1–8.
- Guttman LA. The basis for Scalogram analysis. In: Stouffer SA, Guttman LA, Suchman FA, Lazarsfeld PF, Star SA, Clausen JA, editors. *Studies in social psychology in World War II: Vol 4. Measurement and Prediction*. Princeton: Princeton University Press; 1950. p. 60–90.
- Hagemans ML, Winkel LP, Van Doorn PA, Hop WJ, Loonen MC, Reuser AJ, *et al.* Clinical manifestation and natural course of late-onset Pompe's disease in 54 Dutch patients. *Brain* 2005; 128 (Pt 3): 671–7.
- Hermans MC, Faber CG, De Baets MH, de Die-Smulders CE, Merkies IS. Rasch-built myotonic dystrophy type 1 activity and participation scale (DM1-Activ). *Neuromuscul Disord* 2010; 20: 310–8.
- Hirschhorn R, Reuser AJ. Glycogen storage disease type II : acid a-glucosidase (acid maltase) deficiency. In: Scriver CR, Beaudet AL, Sly W, Valle D, editors. *The metabolic and molecular basis of inherited disease*. 8th edn. New York: McGraw-Hill; 2001. p. 3389–420.
- Hughes RA, Donofrio P, Bril V, Dalakas MC, Deng C, Hanna K, *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: 136–44.
- Kleyweg RP, van der Meche FG, Schmitz PI. Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barre syndrome. *Muscle Nerve* 1991; 14: 1103–9.
- Liang MH. Evaluating measurement responsiveness. *J Rheumatol* 1995; 22: 1191–2.
- Linacre JM. Sample Size and Item Calibration Stability. *Rasch Meas Trans* 1994; 7: 328.
- Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002; 3: 85–106.
- MacAvoy MC, Green DP. Critical reappraisal of Medical Research Council muscle testing for elbow flexion. *J Hand Surg Am* 2007; 32: 149–53.
- Medical Research Council. *Aids to the investigation of the peripheral nervous system*. London: Her Majesty's Stationary Office; 1943.
- Merkies ISJ. Evaluation of scales and measurement instruments in immune-mediated polyneuropathies. Thesis. Rotterdam: Erasmus Medical Center; 2001.
- Merlini L. Measuring muscle strength in clinical trials. *Lancet Neurol* 2010; 9: 1146; author reply 46–7.
- Miescher GC, Steck AJ. Paraproteinaemic neuropathies. *Baillieres Clin Neurol* 1996; 5: 219–32.
- Mitchell SW, Lewis MJ. Tendon-jerk and muscle-jerk in disease and especially in posterior sclerosis. *Am J Med Sci* 1886; 92: 363–72.
- Pallant JF, Tennant A. 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): 1–18.
- Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. 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: 1627–34.
- Prior TW. Technical standards and guidelines for myotonic dystrophy type 1 testing. *Genet Med* 2009; 11: 552–5.
- Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.
- Schreuders TA, Selles RW, Roebroek ME, Stam HJ. Strength measurements of the intrinsic hand muscles: a review of the development and evaluation of the Rotterdam intrinsic hand myometer. *J Hand Ther* 2006; 19: 393–401; quiz 02.
- Shaw F, Wright B, Linacre JM. Disordered steps? *Rasch Meas Trans* 1992; 6: 225.
- Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002; 3: 205–31.
- Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. *J Rehabil Med* 2001; 33: 47–8.
- Tennant A, Conaghan PG. 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: 1358–62.
- Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 2004a; 7 (Suppl 1): S22–6.
- Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004b; 42 (1 Suppl): I37–48.
- The Dutch Guillain-Barré syndrome study group. Treatment of Guillain-Barre syndrome with high-dose immune globulins combined with methylprednisolone: a pilot study. *The Dutch Guillain-Barre Study Group. Ann Neurol* 1994; 35: 749–52.
- Van Asseldonk JT, Franssen H, Van den Berg-Vos RM, Wokke JH, Van den Berg LH. Multifocal motor neuropathy. *Lancet Neurol* 2005; 4: 309–19.
- Van den Berg-Vos RM, Franssen H, Wokke JH, Van den Berg LH. Multifocal motor neuropathy: long-term clinical and electrophysiological assessment of intravenous immunoglobulin maintenance treatment. *Brain* 2002; 125 (Pt 8): 1875–86.
- van der Kooij AJ, Barth PG, Busch HF, de Haan R, Ginjaar HB, van Essen AJ, et al. The clinical spectrum of limb girdle muscular dystrophy. A survey in The Netherlands. *Brain* 1996; 119 (Pt 5): 1471–80.
- van der Meche FG, Schmitz PI. A randomized trial comparing intravenous immune globulin and plasma exchange in Guillain-Barre syndrome. *Dutch Guillain-Barre Study Group. N Engl J Med* 1992; 326: 1123–9.
- van Koningsveld R, Schmitz PI, Meche FG, Visser LH, Meulstee J, van Doorn PA. Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barre syndrome: randomised trial. *Lancet* 2004; 363: 192–6.
- van Nes SI, Vanhoutte EK, van Doorn PA, Hermans M, Bakkers M, Kuitwaard K, et al. Explaining Rasch to neurologists. *Neurology* 2011; 76: 337–45.
- Vandervelde L, Van den Bergh PY, Goemans N, Thonnard JL. ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders. *Neuromuscul Disord* 2007; 17: 459–69.
- Wright BD. Common Sense for Measurement. *Rasch Meas Trans* 1999; 13: 704.