

Published in final edited form as:

J Peripher Nerv Syst. 2014 September ; 19(3): 192–196. doi:10.1111/jns.12084.

Psychometrics evaluation of Charcot-Marie-Tooth Neuropathy Score (CMTNSv2) second version, using Rasch analysis

Reza Sadjadi¹, Mary M. Reilly², Michael E. Shy³, Davide Pareyson⁴, Matilde Laura², Sinead Murphy⁵, Shawna M.E. Feely³, Tiffany Grider³, Chelsea Bacon³, Giuseppe Piscosquito⁴, Daniela Calabrese⁴, and Ted M. Burns¹

¹Department of Neurology, University of Virginia, Charlottesville, VA, USA ²MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, London, UK ³Department of Neurology, College of Medicine, University of Iowa, Iowa City, IA USA ⁴Clinic of Central and Peripheral Degenerative Neuropathies Unit, Department of Clinical Neurosciences, IRCCS Foundation, C. Besta Neurological Institute, Milan, Italy ⁵Department of Neurology, Adelaide & Meath Hospitals incorporating the National Children's Hospital, Tallaght, Dublin, and Trinity College Dublin, Ireland

Abstract

Charcot-Marie-Tooth Neuropathy Score second version (CMTNSv2) is a validated clinical outcome measure developed for use in clinical trials to monitor disease impairment and progression in affected CMT patients. Currently, all items of CMTNSv2 have identical contribution to the total score. We used Rasch analysis to further explore psychometric properties of CMTNSv2, and in particular, category response functioning and their weight on the overall disease progression. Weighted category responses represent a more accurate estimate of actual values measuring disease severity and therefore could potentially be used in improving the current version.

Keywords

Charcot-Marie-Tooth; Charcot-Marie-Tooth Neuropathy Score (CMTNS2) second version; outcome measures; psychometrics; Rasch analysis

Introduction

Charcot-Marie-Tooth disease is the most common type of inherited neuropathy (*Skre, 1974*). Most cases have a length-dependent neuropathy commonly presenting with distal weakness, sensory loss, pes cavus and absent ankle reflexes in the first two decades of life. The CMT Neuropathy Score (CMTNS) was developed to quantify impairment and measure progression in CMT (*Shy et al., 2005*), and was modified from the total neuropathy score (TNS) that was developed to measure impairment and progression in length-dependent sensory neuropathies (*Cornblath et al., 1999*). The CMTNS demonstrated progression of

impairment in patients with CMT1A (*Shy et al., 2008*) and CMT1X (*Shy et al., 2007*). However, the CMTNS did not demonstrate progression in either treated or placebo patients over two years in two large clinical trials of ascorbic acid treatment of CMT1A (*Pareyson et al., 2011; Lewis et al., 2013*). These results suggested that the CMTNS was not sensitive enough to detect change in clinical trials that ran for less than two years.

An international workshop on outcome measures was therefore convened to develop more sensitive outcome measures for CMT (*Reilly et al., 2010*). The CMTNS Version 2 (CMTNSv2) (*Murphy et al., 2011*) resulted from this meeting and was designed to reduce floor and ceiling effects. However, it remained to be shown whether the CMTNSv2 captured different levels of impairment equally and it has not yet been tested in longitudinal studies. It was also not known how the various components of CMTNSv2 related to each other. For example, were CMAP amplitudes and motor testing measuring the same thing or were they independent measures of impairment?

Multiple reviews have tried to explore the methodological limitations of rating scales with ordinal scales such as CMTNS with special emphasis on modern psychometrics, such as item response theory (*Hobart et al., 2007; Cano and Hobart, 2008*). We therefore used Rasch analysis to further evaluate and improve the psychometric properties of CMTNSv2 and its compliance with uni-dimensionality; i.e., assurance that all items were measuring the same “construct” or “concept” (i.e., disease severity in CMTNSv2) (*Rasch, 1980*). The model compares response probabilities for any person attempting different items, measuring whether actual item and person performances are close enough (Item Fitting) to be considered a linear scale (*Bond and Fox, 2007*). Rasch model analysis can help clinicians understand factors contributing to non-linearity of existing scales, and help construct better outcome measures. This information can also offer ideas about modifying scales in order to improve their performance. The major aim of our study was to use Rasch analysis to evaluate the CMTNSv2 on one cohort, comprised of clinical data from 3 international centers and discuss potential changes to ensure that we were capturing a wide range of impairment ranging from mildly to severely impaired. Without this capability we risk being unable to detect small changes in impairment in future natural history studies and clinical trials.

Materials and Methods

Rasch analysis was applied on CMTNSv2 data collected from the centers involved in development of the original outcome measure in the US, the UK and Italy, using Winstep Rasch analysis software version 3.69. Numbers '9 software and Microsoft Excel 2010 were used to further explore data. We tested CMTNSv2 for: 1) Item-person targeting, 2) Item fitting and dimensionality, and 3) Response weighting. Dimensionality test was performed using Principle Component Analysis to measure if minor item or person misfits could potentially form a sub-dimension. Rasch-predicted category responses were used to propose modified category responses to improve overall measuring qualities of CMTNSv2. We chose to focus on patients with CMT1A because future longitudinal impairment studies will likely focus on CMT1A and because patients with CMT1A have the same genetic cause

minimizing the possibility that phenotypic differences between different genotypes might influence our results.

Results

A total of 153 CMTNSv2 completed forms were included from three participating centers (United Kingdom 65, United States 72, Italy 16 CMT1A patients). Overall, there was 84% person and 99% item reliability.

Item-person targeting

“Motor symptoms (arms)” and “Strength (arms)” were more suitable for differentiating disease severity in more disabled patients. “Radial SAP” had more probability of being scored by less disabled patients, thus more suitable for differentiating patients with lower levels of disability. Comparison of item and person distribution on a common logarithmic scale revealed a significant, but modest floor effect, suggesting that items were likely more suitable for moderate to severe forms of disease, with the exception of “Radial SAP” which was suitable for less disability range. This suggested that modifying items to cover less disability range may improve this deficiency. “Pinprick sensibility” and “Ulnar CMAP” were also amongst items more suitable for less disability but did not cover the gap in severity distribution coverage (Table 1; Fig. 1, vertical axis).

Item fitting and dimensionality

There was no major mis-fitting item in the test (Fig. 1, horizontal axis). Universally, all items had good fitting with mean of the squared residuals ranging between 0.83 (“Strength (arms)”) and slightly outfitting 1.45 (“Sensory symptoms”) (Table 1). This indicates that the items belong in the scale and contribute to the overall score of disability.

Forty percent of the total variance in person-ability could not be explained by items. Very subtle misfits from “Sensory symptoms” (1.45), and to a much lesser extent, “Pinprick sensibility” (1.03) could only explain 10% of the total unexplained variance, suggesting that there was a fairly negligible dimensionality effect that should be evaluated further, especially if more sensory items would be added to this outcome in future.

There were “small” and “medium” correlations between most item difficulty measures except for borderline strong correlation between “Sensory symptoms” and “Strength (legs)” (ρ 0.51) (Table 3). There were no significant correlations between “Motor symptoms (legs)”, “Strength (legs)” (ρ 0.19) and “Motor symptoms (arms)”, “Strength (arms)” (ρ 0.10).

Category responses

Category responses for most items were in acceptable order (i.e., the probability of higher response category to be selected by a more disabled patient is higher). The one exception was “Radial SAP,” where more disabled patients scored 2 more often than 3 or 4. This was also seen to a lesser extent in “Strength (arms)” (Table 2).

Category responses for “Motor symptoms (legs)”, “Motor symptoms (arms)” and “Pinprick sensibility” were examples of excellent category response weighting (Table 2). For example, a patient who scored 2 on “Motor symptoms (legs)”, tended to be twice more disabled than a patient who scored 1 on the same item. On the other hand, patients who scored 2 on the “Sensory symptoms” were not necessarily twice as disabled than those scoring 1. This, however, did not create a major misfit in overall performance of CMTNSv2. Also, category responses were disorganized for “Radial SAP” and “Strength (arms)” (Table 2). For example, patients who were more disabled had a higher likelihood of scoring 1 (10-14.9 microV) than 2 (5-9.9 microV). Thus, these category responses, as currently written in the CMTNSv2, will not be able to distinguish different levels of disability. Therefore combining 1 and 2, or even 2 and 3, would potentially improve the overall item fit. A similar concept applies to “Strength (arms)”. For example, category response 3 (<5 on wrist extensors) is more likely to be scored by more disabled patients than 4 (weak above elbow).

In order to improve category response behavior, category responses with similar or disordered range were amalgamated. Rasch-estimated category response weights were sometimes modified (Table 2) and also rounded such that the maximum score would be 40 (Table 2). Reapplying Rasch Analysis using the new amalgamated scale on the very same dataset ameliorated these item properties and also resulted in less of a floor effect, indicating that items will better cover milder range of disease severity (Fig. 2). Thus, we offer a tentative re-working and re-scoring of the response categories (tentatively called CMTNSv2-R). Further analysis of other cohorts may lead to further modification.

Discussion

The CMTNSv2 was designed to increase its sensitivity to detect small differences in impairment that would have been missed by the original CMTNS. As such, we hoped that it would prove to be a more sensitive instrument to detect change in longitudinal studies for slowly progressive disorders such as CMT1A (*Reilly et al., 2010*). Our present study using Rasch analysis of the CMTNSv2, however, demonstrated that in its present form the CMTNSv2 tends to clump impairment scores from many patients in the middle range of severity, though it does distinguish between mildly and severely impaired patients as well. This raises the concern that it will be difficult for the CMTNSv2 to detect small changes in progression that occur in patients who fall between mildly and moderately affected or between moderately and severely impaired as defined by the score. The weighting of responses in the modified scale, shown in Figure 2, suggests an approach in which the scoring is more linear so that smaller differences of clinical change may be more easily detected, at least based on the data obtained during this study. Whether the weighted scale ultimately proves more sensitive to change over time than the present CMTNSv2 will require longitudinal testing. Because the weighted CMTNSv2 was simply calculated from existing scores, these longitudinal studies should be able to be completed in the near rather than the distant future.

We recognize that there are limitations with the CMTNSv2, even in its weighted form. We would argue, in fact, that there is no current “perfect” outcome instrument for CMT, or any chronic neuromuscular disease, and that the design and development of better outcome

instruments for slowly progressing chronic diseases is a science in its own right. Clinician scientists and not patients have developed the CMTNSv2; therefore it is not a patient reported outcome (PRO) instrument. In recognition of the importance of patient input (*Bren, 2006*), we are also in the process of developing and testing patient-reported outcome (PRO) instruments for CMT. These PRO measures are more subjective and therefore unlikely to detect subclinical changes. Some forms of CMT may progress so slowly that the disease may have progressed subclinically, including biologically (demyelination or axonal degeneration), but not altered the patient's life in noticeable ways. If medications are developed to slow demyelination or axonal degeneration, it is important and necessary to know if scientists are "on the right track" biologically even if the patients do not notice benefits, especially in more indolently progressive forms of disease where small changes in one- or two-year clinical trials are anticipated. PROs are also likely affected by psychological cofactors such as patient mood, expectations and response shift, which remain major limitations of PROs for slowly progressive disorders (*Campbell, 1976*). In addition, PRO instruments may provide different answers depending on items such as the age or level of physical impairment of the patients; what is deemed meaningful to a 30 year old may be different than what is meaningful to a 60 year old. Composite outcome measures like the CMTNSv2 also have a place in these studies if they can be shown to be sensitive to change and model the biological underpinnings of the disease process. The CMTNSv2 may not prove to be the perfect Rasch-designed outcome instrument even with its current modifications; however, it is widely utilized to measure impairment in patients with CMT, is easy and efficient, and we believe that further improvement in this scale is valuable, in part because it remains essential to obtain natural history data; thus, we think the CMTNSv2 should be designed as effectively as possible, including having the capability to detect a wide range of impairment in CMT. The CMTNSv2 can then be tested against other and newer instruments to ensure that the most sensitive and meaningful outcome instruments are used in natural history and therapeutic trials.

Acknowledgements

We thank the patients who graciously participated in the study. The project received support from the National Institutes of Health and Office of Rare Diseases (U54NS065712), the Muscular Dystrophy Association (MDA) and the Charcot-Marie-Tooth Association (CMTA).

References

- Bond, TG.; Fox, CM. Applying the Rasch model: fundamental measurement in the human sciences. Lawrence Erlbaum Associates Publishers; Mahwah NJ (Ed): 2007.
- Bren L. The importance of patient-reported outcomes&Its all about patients. *FDA Consum.* 2006; 40:26–32. [PubMed: 17333563]
- Campbell A. Subjective measures of well-being. *Am Psychol.* 1976; 31:117–124. [PubMed: 1267244]
- Cano SJ, Hobart JC. Watch out, watch out, the FDA are about. *Dev Med Child Neurol.* 2008; 50:408–409. [PubMed: 18489458]
- Cornblath DR, Chaudhry V, Carter K, Lee D, Seysedadr M, Miernicki M, Joh T. Total neuropathy score: validation and reliability study. *Neurology.* 1999; 53:1660–1664. [PubMed: 10563609]
- Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol.* 2007; 6:1094–1105. [PubMed: 18031706]

- Lewis RA, McDermott MP, Herrmann DN, Hoke A, Clawson LL, Siskind C, Feely SM, Miller LJ, Barohn RJ, Smith P, Luebke E, Wu X, Shy ME. High-dosage ascorbic acid treatment in Charcot-Marie-Tooth disease type 1A: results of a randomized, double-masked, controlled trial. *JAMA Neurol.* 2013; 70:981–987. [PubMed: 23797954]
- Murphy SM, Herrmann DN, McDermott MP, Scherer SS, Shy ME, Reilly MM, Pareyson D. Reliability of the CMT neuropathy score (second version) in Charcot-Marie-Tooth disease. *J Peripher Nerv Syst.* 2011; 16:191–198. [PubMed: 22003934]
- Pareyson D, Reilly MM, Schenone A, Fabrizi GM, Cavallaro T, Santoro L, Vita G, Quattrone A, Padua L, Gemignani F, Visioli F, Laura M, Radice D, Calabrese D, Hughes RA, Solari A, CMT-TRIAAL; CMT-TRAUK groups. Ascorbic acid in Charcot-Marie-Tooth disease type 1A (CMT-TRIAAL and CMT-TRAUK): a double-blind randomised trial. *Lancet Neurology.* 2011; 10:320–328.
- Rasch, G. Probabilistic models for some intelligence and attainment tests. University of Chicago Press; Chicago: 1980.
- Reilly MM, Shy ME, Muntoni F, Pareyson D. 168th ENMC International Workshop: outcome measures and clinical trials in Charcot-Marie-Tooth disease (CMT). *Neuromuscul Disord.* 2010; 20:839–846. [PubMed: 20850975]
- Shy ME, Blake J, Krajewski K, Fuerst DR, Laura M, Hahn AF, Li J, Lewis RA, Reilly M. Reliability and validity of the CMT neuropathy score as a measure of disability. *Neurology.* 2005; 64:1209–1214. [PubMed: 15824348]
- Shy ME, Siskind C, Swan ER, Krajewski KM, Doherty T, Fuerst DR, Ainsworth PJ, Lewis RA, Scherer SS, Hahn AF. CMT1X phenotypes represent loss of GJB1 gene function. *Neurology.* 2007; 68:849–855. [PubMed: 17353473]
- Shy ME, Chen L, Swan ER, Taube R, Krajewski KM, Herrmann D, Lewis RA, McDermott MP. Neuropathy progression in Charcot-Marie-Tooth disease type 1A. *Neurology.* 2008; 70:378–383. [PubMed: 18227419]
- Skre H. Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clin Genet.* 1974; 6:98–118. [PubMed: 4430158]

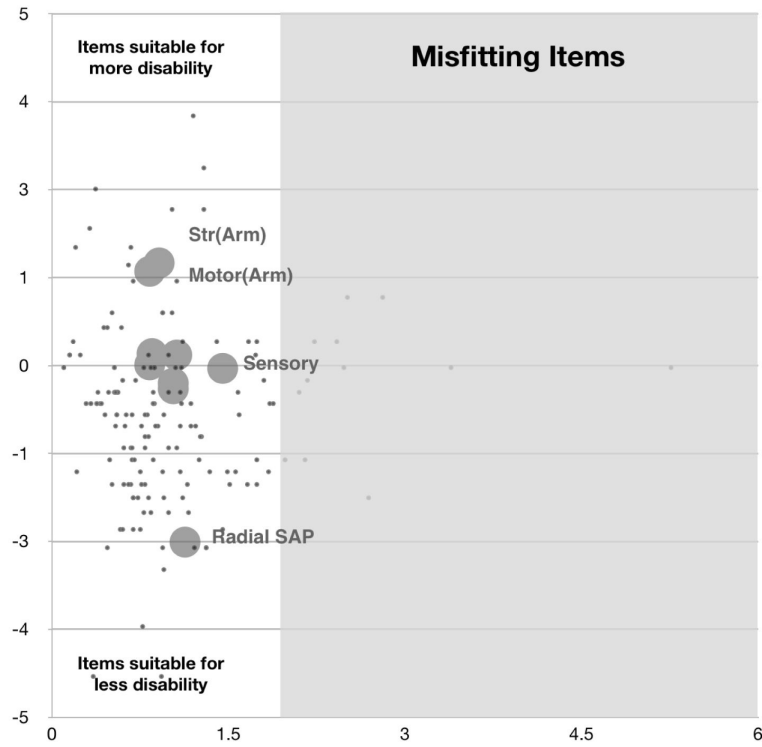


Figure 1.

(Using original scale scoring system)

Vertical axis, represents item difficulty. More difficult items (large circles) are more appropriate for more disabled patients (black dots); Horizontal axis, represents item fit. Items (large circles) and persons (black dots) to the right of vertical line in the grey area are mis-fitting.

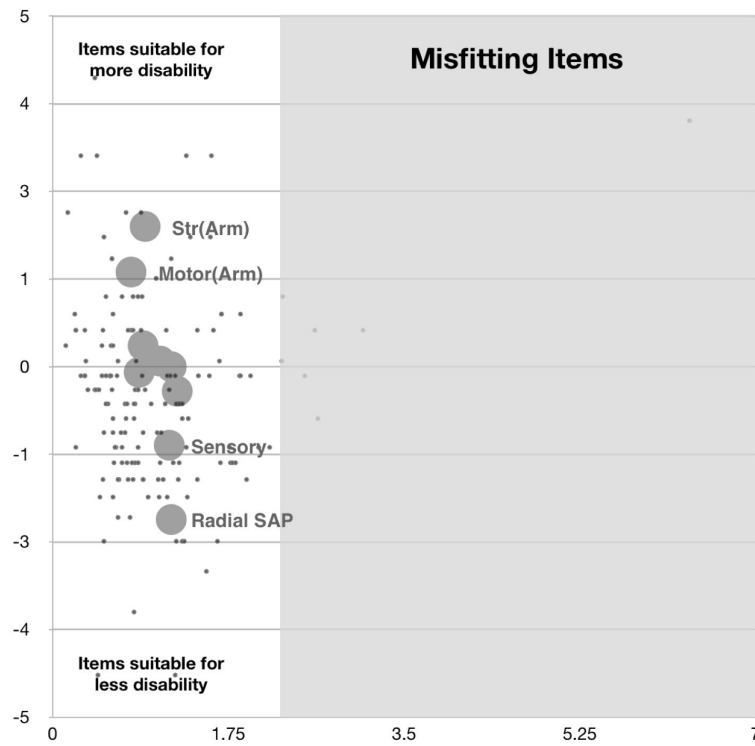


Figure 2.

(Modified Scale, using new scoring system)

Vertical axis, represents item difficulty. More difficult items (large circles) are more appropriate for more disabled patients (black dots); Horizontal axis, represents item fit. Items (large circles) and persons (black dots) to the right of vertical line in the grey area are mis-fitting.

Table 1

Item fit and measure summary; item measure, mean of squared residuals (MSNQ), Outfit Z-score, point-measure correlation (PMC).

Items	Measure	MSNQ*	Z-Score †
Motor symptoms (arms)	1.46	0.91	-0.6
Strength (arms)	1.34	0.83	-1.5
Strength (legs)	0.17	0.85	-1.2
Vibration	0.15	1.06	0.4
Motor symptoms (legs)	0.01	0.83	-1.6
Sensory symptoms	-0.04	1.46	2.8
Ulnar CMAP	-0.25	1.03	0.3
Pinprick sensibility	-0.33	1.03	0.3
Radial SAP	-2.51	1.13	0.5

* Mean of the squared residuals, which represents the unstandardized form of fit statistics.

† Standardized t-value of squared residuals.

Table 2

Item category response (frequencies, measures, original and modified scales).

Items	Category responses	Frequency (%)	Measure	Modified Scoring
Motor symptoms (arms)	0	41	-1.5	0
	1	37	-0.6	2
	2	19	0.4	3
	3	3 †	1.2	5
	4	1 †	2.2	6
Strength (arms)	0	38	-1.5	0
	1	44	-0.7	2
	2	15	0.7	4
	3	1 †	2.7 *	5
	4	1 †	2.3	5
Strength (legs)	0	25	-1.9	0
	1	39	-0.9	1
	2	22	0.1	2
	3	8	0.1	3
	4	6	1.7	3
Sensory symptoms	0	31	-1.7	0
	1	15	-0.6	1
	2	26	-0.6 *	1
	3	22	0	3
	4	7	0.9	3
Vibration	0	20	-2.1	0
	1	21	-1.0	2
	2	14	-0.5	2
	3	43	0	3
	4	1 †	2.0	3
Motor symptoms (legs)	0	7	-2.4	0
	1	37	-1.3	1
	2	42	-0.4	3
	3	12	0.9	5
	4	2 †	1.9	6
Ulnar CMAP	0	18	-1.8	0
	1	28	-1.2	2
	2	31	-0.6	3
	3	15	0	4

Items	Category responses	Frequency (%)	Measure	Modified Scoring
	4	8	1.9	5
Pinprick sensibility	0	15	-1.9	0
	1	31	-1.3	2
	2	22	-0.5	3
	3	25	0.1	4
	4	7	1.2	5
Radial SAP	0	3 †	-3.6	0
	1	3 †	-1.4	2
	2	16	-1.7 *	3
	3	14	-1.1	3
	4	65	-0.3	4

* Average measure does not ascend with category score.

† Item frequencies are relatively small.

Table 3

Item dependencies.

	Sensory	Motor (Leg)	Motor (Arm)	Pinprick
Motor (Leg)	-0.03			
Motor (Arm)	-0.16	-0.10		
Pin	0.33	-0.18	-0.11	
Vibration	-0.24	-0.16	-0.11	0
Strength (Leg)	-0.51	0.19	-0.12	-0.40
Strength (Arm)	-0.47	-0.18	0.10	-0.20
Ulnar CMAP	-0.24	-0.17	-0.09	-0.45
Radial SAP	-0.17	-0.16	-0.13	-0.21

	Vibration	Strength (Leg)	Strength (Arm)	Ulnar CMAP
Strength (Leg)	0.09			
Strength (Arm)	0.03	0.22		
Ulnar CMAP	-0.28	-0.06	0.10	
Radial SAP	-0.28	-0.05	-0.18	0.20