Peripheral neuropathy outcome measures standardisation (PeriNomS) study part 2 : getting consensus

Valorisation-addendum

Valorisation in scientific research describes the value added to the society. Besides education and scientific research, valorisation is one of the key tasks of universities. As such, it is important to explicitly state the added value of the research performed.

1 Social and economic relevance?

Clinical trials in patients with immune-mediated neuropathies like Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), and monoclonal gammopathy of undetermined significance related neuropathy (MGUSP) often used different outcome measures to present their results. In the past, multiple types of often ordinal outcome measures have been used in these disorders leading to different results of studies. The design and validation process of creating an outcome measure is a time consuming process with a burden to patients (they have to complete the same outcome measure several times) and with considerable costs. Furthermore, in order to obtain comparability between trials, researchers should strive to use the same existing outcome measures.

Proper attention should be given to the choice of the best outcome measure, particularly choosing outcome measures that have been tested in terms of their clinimetric properties in the disease of interest. The use of insensitive outcome measures might be the reason for trials in the inflammatory neuropathy field being negative. This could mean that the Food and Drug Administration (FDA), due to the use of improper outcome measures, is not approving treatments that might have an effect, which may hurt patients.

Most outcome measures used thus far in inflammatory neuropathies are at the ordinal level and based on the classical test theory. The Rasch model overcomes the disadvantages of ordinal-based outcome measures. We constructed new outcome measures and evaluation new and existing outcome measures using the Rasch method. These outcome measures could overcome the previously presented problems.

In addition, trial results evaluating treatment options are often driven by the presence or absence of having a significant (p-value) difference between the various (e.g. treated versus placebo) groups. However, the presence of a statistical significant difference does not always mean that the findings are clinically relevant. The minimum clinically important difference (MCID) was introduced in the current thesis as the minimum change in score necessary to reflect a clinically relevant change.[1] There are several methods available to calculate the MCID, thus showing its many faces and the lack of international consensus on this matter.[2] The biggest disadvantages of applying these MCID techniques is perhaps the standard error (SE) being considered "static (unchanged)"

throughout the range of the outcome measure being used. The MCID-SE concept using Rasch-transformed demonstrates a dynamic pattern of 'being-a-responder' in patients with immune-mediated neuropathies. SE values are the lowest in the middle part of the metric and increases towards the edges of the metric's range. A responder is defined if a patient showed a significant change exceeding MCID-SE ≥ 1.96.
Using the concept of MCID and responsiveness at the individual level could give physicians the opportunity to determine scientifically and objectively whether the patient is a responder or not. When the patient is not responding, one may consider to discontinue the treatment or to change to different treatment options. This would contribute to personalized medicine and could, at the same time, lower the costs in health care.

2 Target groups

Outcome measures are usually developed using the Classical Test Theory (CTT). However, outcome measures based on CTT may constitute items that are arbitrarily collected with ordinal response options (e.g. 0 never, 1 seldom, 2 quite often, 3 very often, 4 always).
Although the distance between the different response options appear linear (score 1,2,3,4) is the true distance between the different categories not know and most probably unequal. Physicians often consider as an example a 1-point response change for an item (e.g., from 0 to 1) equivalent to a 1-point change from 2 to 3.
Also, patients are requested to complete all items, even though some may be irrelevant or inappropriate for their level of ability. A sumscore of the scale's items is often calculated 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. Based on these shortcomings, CTT-based outcome measures may limit the comparison of patients and study results.
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 instead of ordinal scores would give a true reflection of disease impact, of differences between individuals and groups, and of treatment effects. One of the widely used approaches is the Rasch method, which was introduced by the Danish mathematician Georg Rasch. 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. Rasch analysis transforms obtained ordinal scores into interval measures and places both items and patients' parameter estimates on the same logodds units (logit) scale. 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. In order to create an interval outcome measure using the Rasch model, several criteria have to be fulfilled. RUMM2030 is a computer program that 'checks' whether the raw data fulfil all these criteria. There are several ways to deal with items deviating from the models' expectations.
We have used a modern scientific approach, based on the Rasch model, leading to disease-specific, true interval measurement, and therefore enable highly improved evaluation and construction of outcome measures in interventional trials. There are several pharmaceutical companies, like Baxter, Talecris, Octapharma that are involved in development of immunomodulating therapy. Although creating outcome measures using the Rasch model is not novel, the outcome measures developed by our group are unique and novel. The advantages are that we provide an interval outcome measure for specific diseases. Moreover, using these outcome measures provides the opportunity to compare the results of different research or trials on the same linear ruler, thereby improving comparability of different studies.

Muscle strength testing is frequently used in clinical trials as a primary outcome measure at the impairment level and can be measured using a tool like the Marin Vigorimeter [3] or by manual muscle testing (Medical Research Council (MRC) grades.[4] The MRC is the most widely applied tool to manually measure strength of muscle groups at bedside.[5] We demonstrated that physicians, independent of their experience or the type of neuromuscular illness examined, are unable to apply the MRC grades in a proper manner.[6] Also, the MRC grading system failed to differentiate various degrees of muscle weakness. In order to solve these problems, the MRC grades were modified to a homogenous four category response options. The four response categories are still ordinal based; therefore meaningful sum scores can only be made after transforming the data through Rasch. After Rasch modelling, we were able to present a transformed modified MRC 12 muscle groups summed score for GBS and CIDP. The Vigorimeter demonstrated the ability to capture clinically meaningful changes quite early in the interventional phase when compared with the primary ordinal-based outcome measure used.[7, 8]
Although the MRC grading system is widely used, it holds several disadvantages. The Vigorimeter overcomes these advantages, however this tool only measures grip strength. These results might trigger companies to develop devices to measure strength at the interval levels. These devices should be, easy applicable, validated and not too expensive. Such a tool could also be used in clinical practice for the follow-up of patients.

3 Activities and products

Creating and (re)analyse outcome measures to develop a Rasch-based tool at the interval level is a time-consuming and labour-intensive process, that warrants specific

expertise. The industry may prefer to buy our Rasch-build outcome measures rather than developing it themselves. Through paid licensing, we can offer companies (and other academia) our developed Rasch-build outcome measures including the required analysis key. Analysing the results obtained from Rasch-build outcome measures requires expertise. It takes several years before that expertise is obtained, and therefore we are able to offer this service to companies and other universities.

4 Innovation

Existing tools for generating outcome measures are usually based on Classical Test Theory (CTT). CTT method holds several disadvantages, thereby creating an illusion of interval measurement. Therefore, CTT-based outcome measures limit the comparison of patients and study results. We have used a modern scientific approach, based on the Rasch model, leading to disease-specific, true interval measurement, and therefore highly improved evaluation and construction of outcome measures in interventional trials. Creating outcome measures using the Rasch model is not novel. However, the outcome measures developed by our group are unique and novel. The advantages are that we provide an interval outcome measure for specific diseases that are validated and responsive. Moreover, using these outcome measures provides the opportunity to compare the results of different research or trials on the same linear ruler, thereby improving comparability of different studies.

5 Implementation

We offer our outcome measures for use to companies and academia, through paid licensing. Moreover, we are able to perform the required analysis of the obtained data. These tools are intended to be used in clinical trials. For obtaining a license, please contact dr. Ingemar S.J. Merkies or prof.dr. C.G. Faber.

Literature

1. 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.
2. 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.* Current opinion in rheumatology, 2002. **14**(2): p. 109-14.
3. 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.
4. Medical Research Counsil, *Medical Research Council. Aids to the investigation of the peripheral nervous system.* 1943, London: Her Majesty's Stationary Office.
5. Dyck, P.J., et al., *History of standard scoring, notation, and summation of neuromuscular signs. A current survey and recommendation.* Journal of the peripheral nervous system : JPNS, 2005. **10**(2): p. 158-73.
6. Vanhoutte, E.K., et al., *Modifying the Medical Research Council grading system through Rasch analyses.* Brain : a journal of neurology, 2012. **135**(Pt 5): p. 1639-49.
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 neurology, 2008. **7**(2): p. 136-44.
8. Vanhoutte, E.K., et al., *Vigorimeter grip strength in CIDP: a responsive tool that rapidly measures the effect of IVIG--the ICE study.* Eur J Neurol, 2013. **20**(5): p. 748-55.