Global PROMS Initiative data analysis workshop

Tuesday, 11 November 2025 (14:00–18:30)

David Andrich* **(Workshop Leader)** and **Scientific Planning Committee**
(Jeremy Hobart (Chair)*[2], **Maria Pia Sormani***[3], **Gary Cutter***[4] and **Tomas Kalincik***[5])

Year on year, PROMs play increasingly important and influential roles in MS clinical trials, studies and practice. We hope and expect that the PROMs we choose will produce "reliable and valid" scores; that is, adequate estimates of the clinical variables they intend to measure, so that our analyses of these estimates produce results we can trust.

This interaction, example and discussion-based workshop concerns this issue.

There are different approaches to determining if PROMs produce scientifically adequate estimates. The widely used "traditional psychometric methods" are straightforward, but scientifically and clinically very limited. The rarely used "modern" psychometric methods are less straightforward but scientifically profound and extremely informative clinically.

They enable detailed, clinically insightful examinations of PROMs, diagnosing measurement problems and providing paths to their solution. Also, under the right circumstances, they enable the transformation of ordinal-level PROMs scores into interval-level estimates (measurements) with individual-person standard errors, thereby enabling examinations of statistical significance at the individual person level as well as group level.

This workshop concerns the modern psychometric method Rasch measurement theory (RMT).

Whilst many have heard of RMT, or Rasch analysis, few will be familiar with its capabilities and potentials to radically change our PROMs-related work. We will present, and discuss interactively, a series of non-technical, clinically relevant demonstrations of the added value of using RMT routinely in our PROMS work.

Naturally, one-off workshops can only achieve a limited amount, but we hope to raise interest and awareness, and start a programme of, education and application of RMT routinely to MS PROMs related work that can advance measurement methods in clinical trials that can reduce their size, duration and cost whilst increasing their accuracy and quality.

Exemplars, chosen to highlight specific measurement issues, will include some of the following PROMS:
From MS studies: six fatigue scales (MFIS, FSMC, NFI-MS, PROMIS, NeuroQol-19, FSIQ-RMS); ABILHAND; SymptoMScreen; EQ-5D; MSIS-29; MSWS-12. From Cancer studies: EORTC-30

…..and, as RMT can be applied to clinician-rated scales, some of the following clinical scales:
EDSS; Ambulation index; Rankin Scale; Ashworth Scale; Rivermead Mobility Index; ALS-Functional rating scale, Functional independence measure (FIM); Barthel Index, Unified Huntington's Disease Rating Scale, ADAS-cog.

Agenda

Introduction to workshop and Rasch Measurement Theory for clinicians

Time	Duration	Session Title
14:00-14:15	15′	**Welcome** The workshop will consist of **a set of interactive demonstrations of the clinically meaningful advantages of using RMT analyses in PROM work**. The demonstrations will be from clinical trial and clinical study data. **Most of the demonstrations will be from MS studies**; however, we have a few valuable and highly relevant demonstrations from other neurological diseases and cancer studies.

Session 1: **Interactive demonstrations and discussion**

Time	Duration	Title
14:15-16:00	1h45′	**Interval measures rather than ordinal scores** Most PROM data analyses use an instrument's total score, from summing some or all the item scores, as the measure (estimate) of the clinical variable. For example, the MSWS-12 version 1 has 12 questions. Each question has five response categories scored 1-5. The total score is the sum of the item scores and ranges from 12-60. Both item and total scores are recognised to be ordinal level data. In contrast, RMT enables these ordinal scores to be transformed into interval estimates, under appropriate conditions. In this demonstration **we show that the relationship between PROM total scores and interval measures is non-linear and address the implications of this for clinical trials**. We will address briefly how these estimates are generated and the conditions required. **Individual level analyses as well as group-level analyses** Most PROM data analyses using total scores are only amenable to group-level analyses. In clinical trials, individual level change is often benchmarked against an estimate of a "clinically meaningful change" that is typically heuristic. In contrast, RMT generates, for every person's interval-level estimate, a formal standard error derived from the person's responses to the items. These standard errors enable legitimate individual level statistical comparisons. That is, for each individual person in a clinical trial, RMT enables analysts to determine if they had a statistically significant change. In this demonstration, **we show examples of these standard errors, and how this enables a test of significant change at the individual person level, and how these analyses complement group-level analyses.** We also show why and how heuristic interpretations of "clinically" significant change have fundamental flaws. **Visualising PROM instruments and their measurement precision** Most PROM analyses concentrate on the total scores people get. We do not get to "see" the instruments. RMT enables visualisations, and these are clinically informative at multiple levels.

In this demonstration **we show visualisations of a range of PROMs** (and some clinician rated scales), **what clinically useful information these pictures convey, what they mean and their implications for interpreting study results**.

Determining if the response categories work as intended

The response categories of any item of any PROM try to map out more or less of the problem the item addresses. These response categories – e.g., none, mild, moderate, severe - are a set of descriptions of different magnitudes of the item.

These categories must "work" empirically in the order they are intended conceptually so that the total scores produced by summing items are meaningful. This requirement that the item response categories work empirically as intended conceptually becomes particularly important when multiple items are combined to give a single value as it underpins the meaningfulness of the instrument and its scores.

Standard methods of PROM analysis do not enable the workings of response categories to be examined. In contrast, RMT enables a detailed examination of the workings of item response categories.

In this demonstration, which builds on demonstration 3) as it is inherent to the interpretation of instrument visualisations, we show that RMT provides a formal test of the hypothesis that the item response categories work empirically as intended conceptually. We address what the results mean, and what to do, when the item response categories do not work as intended.

16:00-16:30 **Coffee break**

Session 2: **Interactive demonstrations and discussion**

Time	Duration	Title
16:30-18:15	1h45'	**Visualising the alignment between the instrument and the sample**

An important feature of PROMs is that they measure over a fixed, often narrow range. In doing so, the relationship between the instrument and the study sample – in terms of the alignment between the range measured by the scale and the range measures in the sample – affects the interpretation of results. We have not seen this addressed in clinical studies before.

In this example, **we show the alignment between scale and sample for several studies, and how this can affect the interpretation of change, and how misalignment causes misinterpretation of results.**

"Validating" PROMs: moving beyond reliability and correlations between PROMs

Most PROM validation studies involve computing correlations between scores generated by different PROMs and other variables. The general aim being to show higher correlations between like variables than between unlike variables. Whilst this is the standard approach it limited information about exactly what the instrument is measuring (its validity).

RMT enables detailed and forensic examinations of PROM instruments to determine different aspects of their reliability and validity. **This demonstration shows the wealth of information that can be derived from an RMT analysis of PROM data**, how this affects study results, informs instrument development, and enhances our understandings of the clinical variables we are intending to measure.

Comparing instruments head-to-head

Traditional comparisons of PROM measures intended to measure the same variables are limited and indirect. RMT enables a detailed comparison of instruments seeking to measure the same variables.

In this demonstration **we compare different fatigue PROMs to determine the extent to which they measure the same variable, and how the analysis brings out the similarities and differences between the different instruments**.

Reliability is a sample dependent statistic

Studies often report instruments as "reliable and valid" implying one-off examinations are adequate. However, reliability is a sample-dependent statistic and only pertains to a specific PROM in a specific sample.

Traditional methods report PROMS reliability using either an indicator of internal consistency of the items (typically Cronbach's alpha) and/or test-retest reproducibility. In this demonstration we show the reliability of different instruments, discuss how it is computed and its interpretation, show Cronbach's alpha is frequently inflated, and shows that test-retest reproducibility conflated instrument and sample variability.

We show RMT provides a more advanced method of examining reliability and of testing test-retest reproducibility that formally tests the stability of the instrument's performance.

Improving existing PROM instruments

New instruments are rarely revised or advanced. Partly, because traditional methods only give limited information on PROMs.

In this demonstration **we show how an RMT analyses can identify measurement strengths and weaknesses and provide the platform for improving PROM measurement.**

Developing new measures

Bringing all the information together from the above demonstrations we outline an approach to developing new instruments.

Closing: **Turning Insight into Impact**

Time	Duration	Title
18:15-18:30	15′	**What RMT-based work should the MS community and Global PROMS Initiative take on?**

Biographies of Workshop Members

David Andrich*

The team is known to the MS community, except for David Andrich; Emeritus Professor at The University of Western Australia and Fellow of the Academy of Social Sciences of Australia for his contributions to measurement in the social sciences.

Over 5 decades, following a PhD from The University of Chicago and working directly with Georg Rasch, David has developed the RMT field. His knowledge of RMT, its theory and mathematics, applications and interpretations, and measurement in general is unbeatable. He is the author of **Rasch models for measurement**. Sage University Paper Series on Quantitative Applications in the Social Sciences: Sage Publications; Andrich, D. & Marais, I **A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences**. Springer, and many seminal papers in the field, as well as the powerful Software, Rasch Unidimensional Measurement Models (RUMM2030).

Jeremy Hobart*[2]

Consultant Neurologist and Chief Scientific Officer, Transform MS CIC, Hon. Prof of Clinical Neurology, **University of Plymouth**

Gary Cutter*[4]

Senior Research Scientist at IHMC and emeritus professor of biostatistics at the **University of Alabama at Birmingham School of Public Health**

Maria Pia Sormani*[3]

full professor of biostatistics at the **University of Genoa**, known for her significant contributions to multiple sclerosis research

Tomas Kalincik*[5]

Dame Kate Campbell Professorial Fellow, the head of the Clinical Outcomes Research (CORe) Unit at the **University of Melbourne** and of the Neuroimmunology Centre at the Royal Melbourne Hospital