

Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods

J Hobart* and S Cano

Neurological Outcome Measures Unit, Department of Clinical Neurosciences, Peninsula College of Medicine and Dentistry, Plymouth, UK, and Institute of Neurology, London, UK

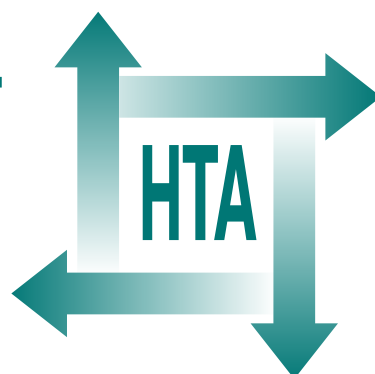
*Corresponding author



Executive summary

Health Technology Assessment 2009; Vol. 13: No. 12
DOI: 10.3310/hta13120

**Health Technology Assessment
NIHR HTA Programme
www.hta.ac.uk**





Executive summary

Background

Rating scales are used increasingly as measurement instruments in clinical trials, clinical studies, clinical audit and clinical practice. The results of these studies influence the care of individual people, the making of health policy and the direction of future research. The inferences made from these studies are based on the analysis of numbers generated by the rating scales they use as outcome measures. If clinically meaningful interpretations are to be made from these studies, it is a requirement that the rating scales used are rigorous measures of the variables (aspects of health) they claim to quantify.

This report concerns psychometric methods: these are methods for developing and evaluating rating scales, and for analysing their data. There are many different psychometric methods for evaluating scales in health measurement. Each uses a different type of evidence to determine the extent to which a scale has achieved its goal of generating measurements. This monograph concerns three psychometric methods: traditional psychometric methods, Rasch measurement and Item Response Theory (IRT).

Objective

We evaluate the added value of the new psychometric methods over existing 'traditional' psychometric methods. The report is in two parts. Chapters 1–3 concern theory. Chapters 4–8 are practical demonstrations using existing sets of rating scale data. The report is aimed at clinicians and researchers working in health measurement and tries to provide clear, detailed, non-technical explanations, and a link into the existing but somewhat inaccessible and abstruse literature. The practical demonstrations are comprehensive with full explanations and extensive visual illustrations. There is repetition across chapters to ensure that the basic principles are conveyed.

Methods

The first part of this monograph (Chapters 1–3) presents reviews of the existing literature. Chapter 1 concerns the role of rating scales and the theory and practice of traditional psychometric methods. Chapter 2 outlines the impetus behind the new psychometric methods (Item Response Theory and Rasch measurement), charts their development, and explains their similarities and differences. In this chapter, we provide the case underpinning the reasons why the rest of the monograph focuses on Rasch measurement and not on Item Response Theory. Chapter 3 describes the theory behind Rasch measurement, the development of the Rasch measurement model, the properties of the model and how it 'works' in practice.

The second part of this monograph (Chapters 4–8) presents five practical head-to-head comparisons of Rasch analysis and traditional psychometric methods based on data sets produced from a variety of settings. Chapter 4 compares evaluations of the Rivermead Mobility Index (RMI) in 666 people with multiple sclerosis (MS). Chapter 5 compares evaluations of the Multiple Sclerosis Impact Scale (MSIS-29) in 1725 people with MS. Chapter 6 compares evaluations of test–retest reliability of the MSIS-29 in 150 people with MS. Chapter 7 demonstrates the use of Rasch measurement to equate four scales measuring physical functioning and four scales measuring psychological functioning. Chapter 8 compares the evaluation of relative responsiveness of the Barthel Index and Functional Independence Measure motor scale in 1400 people admitted to a neurorehabilitation unit.

Results

Our reviews of the health measurement literature reveal that: (1) the dominant traditional paradigm for the construction, evaluation and analysis of scales (traditional psychometric methods) is based

on a weak theory; (2) new psychometric methods (Rasch measurement and Item Response Theory) represent a concerted attempt to bring theory and structure to an inherently weak field; and (3) Rasch measurement and Item Response Theory are fundamentally very different approaches.

In the second half of the monograph we focus on worked examples comparing Rasch measurement with traditional psychometric methods. In Chapters 4 and 5, our comprehensive evaluations of the Rivermead Mobility Index (RMI) and the Multiple Sclerosis Impact Scale (MSIS-29) reveal the limitations of traditional psychometric methods and demonstrate the advantages of Rasch measurement. In Chapter 6 we demonstrate the use of different data designs to answer the various components of complex problems and the examination of differential item functioning in test–retest reliability. In Chapter 7 we demonstrate the use of equating tables that enable users of different scales to compare their results. Finally, in Chapter 8 we find that group-based statistics may mislead, and highlight the value and importance of being able to examine change data at the individual person level.

Conclusions and recommendations

We believe that when taken together the arguments and demonstrations in this monograph, both

theoretical and empirical, illustrate that Rasch measurement is vastly superior to traditional psychometric methods. Although we have highlighted the value of Rasch measurement in the context of only a limited number of scales for people with MS, we feel that it has much to offer all health measurement, state-of-the-art clinical trials and, most importantly, the individual patients treated by clinicians.

There are a number of future research directions. As next steps, we recommend: (1) that other researchers and clinicians reproduce our findings in a range of clinical populations; (2) detailed head-to-head comparisons of Rasch measurement and Item Response Theory; (3) work to determine further sample size requirements for adequate person and item estimations; and (4) exploration of the application of Rasch measurement to clinical practice in areas including prioritising problems, facilitation of communication, screening potential problems, identifying preferences, monitoring changes or responses to treatment, training new staff and clinical audit.

Publication

Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;**13**(12).

NIHR Health Technology Assessment Programme

The Health Technology Assessment (HTA) Programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA Programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA Programme is needed in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA Programme then commissions the research by competitive tender.

Second, the HTA Programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this issue of the journal was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formally transferred to the HTA Programme in April 2007 under the newly established NIHR Methodology Panel. The HTA Programme project number is 95/01/05. The contractual start date was in February 2005. The draft report began editorial review in January 2007 and was accepted for publication in March 2008. The commissioning brief was devised by the NCCRM who specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

Editor-in-Chief:

Professor Tom Walley

Series Editors:

Dr Aileen Clarke, Dr Peter Davidson, Dr Chris Hyde, Dr John Powell,
Dr Rob Riemsma and Professor Ken Stein

ISSN 1366-5278

© 2009 Queen's Printer and Controller of HMSO

This monograph may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Applications for commercial reproduction should be addressed to: NCCHTA, Alpha House, Enterprise Road, Southampton Science Park, Chilworth, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NCCHTA.

Printed on acid-free paper in the UK by the Charlesworth Group.