

- 16 Hagquist C, Malin BM, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud* 2009; **46**: 380–93.
- 17 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9 validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–13.
- 18 Wittkamp KA, Naeijee L, Schene AH, Huyser J, Van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007; **29**: 388–95.
- 19 Gilbody S, Richards D, Brearley S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007; **22**: 1596–1602.
- 20 Masters GA. Rasch model for partial credit scoring. *Psychometrika* 1982; **42**: 149–74.
- 21 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**: 1–18.
- 22 Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psych* 2009; **9**: 21.
- 23 Marais I, Andrich D. Effects of varying magnitude and patterns of response dependency in the unidimensional Rasch model. *J Appl Meas* 2008; **9**: 105–24.
- 24 Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al; PROMIS Cooperative Group. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007; **45** (suppl. 1): S22–31.
- 25 Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002; **3**: 205–31.
- 26 Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care* 2004; **42**: 7–16.



The application of Rasch measurement theory to psychiatric clinical outcomes research

Commentary on . . . Screening for depression in primary care[†]

Skye P. Barbic,¹ Stefan J. Cano²

BJPsych Bulletin (2016), **40**, 243–244, doi: 10.1192/pb.bp.115.052290

¹University of British Columbia, Vancouver, Canada; ²Modus Outcomes, Stotfold, UK

Correspondence to Stefan Cano (stefan.cano@modusoutcomes.com)

First received 10 Aug 2015, accepted 5 Nov 2015

© 2016 The Authors. This is an open-access article published by the Royal College of Psychiatrists and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Summary This commentary argues the importance of robust, meaningful assessment of clinical and functional outcomes in psychiatry. Outcome assessments should be fit for the purpose of measuring relevant concepts of interest in specific clinical settings. As well, the measurement model selected to develop and test assessments can be critical for guiding care. Three types of measurement models are presented: classical test theory, item response theory, and Rasch measurement theory. To optimise current diagnostic and treatment practices in psychiatry, careful consideration of these models is warranted.

Declaration of interest None.

Unlike many fields in medicine, most clinical outcomes in psychiatry are not directly observable and cannot be captured with diagnostic tests such as blood work or imaging. In recent years, the importance of the routine use of clinical outcome assessments (patient-reported outcomes, clinician-reported outcomes, observer-reported outcomes and performance outcomes) for measuring the symptoms of disease and treatment outcomes has been increasingly emphasised.¹ Clinical outcome assessments such as the Patient Health Questionnaire-9 (PHQ-9)² are

now commonly used in clinical research and practice to provide an assessment of a patient's severity of mood and improvement in response to treatment.³ More broadly, as the demand increases for a broad range of mental health services to be patient-centred, clinical outcome assessments are used to capture outcomes such as sustained symptom reduction, return to full functioning and optimal patient well-being.⁴

To optimise mental healthcare, clinical outcome assessments used in psychiatry should be shown to be fit for purpose. They should appropriately capture the concept of interest (e.g. depression) in the context of use (e.g.

[†]See original paper, pp. 237–243, this issue.

patients attending primary care clinics reporting symptoms of depression).¹ They should also be underpinned by an appropriate measurement model, that is they should have evidence that the summed score of their individual items is 'psychometrically sound'.¹ To this end, there are three main psychometric approaches based on three types of measurement model: classical test theory (CTT), Rasch measurement theory (RMT) and item response theory (IRT).⁵

The current dominant paradigm in clinical outcomes research is CTT, the foundations of which were laid down by Charles Spearman at the turn of the twentieth century.⁶ CTT is associated with the psychometric properties most commonly recognised and understood by clinicians (e.g. reliability, validity and ability to detect change). However, there are four important limitations⁷ that prevent CTT methodology from fulfilling the requirements of scientific rigour demanded of high-stakes clinical decision-making: (a) measurements generated are ordinal rather than interval; (b) scores for persons and samples are scale dependent; (c) scale properties, such as reliability and validity, are sample dependent; (d) data can support group-level inferences but are not suitable for individual patient measurement.

Georg Rasch, a Danish mathematician, argued that the core requirement of social measurement should be the same as that in physical measurement, and developed the simple logistic model now known as the 'Rasch model'.⁸ In essence, RMT methods assess the extent to which observed clinical outcome assessment data (e.g. patient ratings on the items of the PHQ-9) 'fit' with predictions of those ratings from the Rasch model (which defines how a set of items should perform to generate reliable and valid measurements).⁸ The difference between the expected and observed scores reveals the extent to which valid measurement is achieved. In turn, this gives rise to a range of potential investigations to better understand the extent to which the clinical outcome assessment under investigation is an appropriate measurement instrument (e.g. scale-to-sample targeting, adequacy of type and kind of response options, item and person fit, item dependency (or bias), stability between sub-groups).^{7,9} Importantly, RMT addresses⁷ each of the four limitations of CTT described above: (a) linear measurements can be constructed from ordinal-level data; (b) item estimates provided are free from the sample distribution and person estimates are free from the scale distribution; (c) subsets of items from each scale rather than all items can be used (i.e. the foundation for item banking and computerised adaptive testing); (d) estimates are suitable for individual person analyses rather than only for group comparison studies.

IRT is another body of psychometric methodology that is used to ascertain the degree to which a given model and parameter estimates can account for the structure of and statistical patterns in a clinical outcome assessment dataset.¹⁰ The distinction between RMT and IRT is subtle but important. IRT models are statistical models used to explain data, and the aim of an IRT analysis is to find the statistical model that best explains the observed data.⁹ By contrast, the aim of RMT is to determine the extent to which observed clinical outcome assessment data satisfy the measurement model.⁸ When the data do not fit the model,

they are examined to try to explain the misfit. This is the central tenet of the Rasch model and one that distinguishes it from IRT models. Specifically, its defining property is its mathematical embodiment of the principle of invariant comparison. Thus, the comparison of two people is independent of which items are used within a set of items assessing the same concept of interest. In this way, the Rasch model is taken as a criterion for the structure of the responses, rather than simply a statistical description of the responses from patients. This central tenet distinguishes the RMT diagnostic paradigm from the IRT modelling paradigm.⁹

In this issue, Horton and Perry provide an example of diagnostic information that can be attained using RMT methods, not available using information gleaned from CTT or IRT methods.¹¹ The availability and increased application of RMT psychometric methods for developing and evaluating clinical outcome assessments in psychiatry has important implications for future research and practice. By better understanding the strengths, weaknesses and measurement potential of such assessments, we are able to build an evidence base towards optimising the organisation and delivery of healthcare in psychiatry.¹²

About the authors

Skye P. Barbic (PhD) is post-doctoral fellow at the Faculty of Medicine, Department of Psychiatry, University of British Columbia, Vancouver, Canada, and **Stefan J. Cano** (PhD, CPsychol, AFBPsS) is co-founder and chief science officer of Modus Outcomes, Stotfold, UK.

References

- 1 US Food and Drug Administration. Clinical Outcome Assessment Qualification Program. FDA, 2015. Available at: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentTools/QualificationProgram/ucm284077.htm> (accessed 23 November 2015).
- 2 Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–13.
- 3 Thase M. Translating clinical science into effective therapies. *J Clin Psychiatry* 2014; **75**: e11.
- 4 Thornicroft G, Slade M. New trends in assessing the outcomes of mental health interventions. *World Psychiatry* 2014; **13**: 118–24.
- 5 Cano S, Hobart J. The problem with health measurement. *Patient Pref Adher* 2011; **5**: 279–90.
- 6 Spearman CE. The proof and measurement of association between two things. *Am J Psychol* 1904; **15**: 72–101.
- 7 Hobart J, Cano S. Improving the evaluation of therapeutic intervention in MS: the role of new psychometric methods. *UK Health Techn Assess Prog (Monograph)* 2009; **13**: 1–200.
- 8 Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Education Research, reprinted: MESA Press, 1993.
- 9 Andrich D. Rating scales and Rasch measurement. *Expert Rev. Pharmacoecon Outcomes Res* 2011; **11**: 571–5.
- 10 Lord FM, Novick MR. *Statistical Theories of Mental Test Scores*. Addison-Wesley, 1968.
- 11 Horton M, Perry A. Screening for depression in primary care: a Rasch analysis of the PHQ-9. *BJPsych Bull* 2016; doi: 10.1192/pb.bp.114.050294.
- 12 Barbic S, Kidd S, Davidson L, McKenzie K, O'Connell M. Validation of the Brief Version of the Recovery Self-Assessment (RSA-B) using Rasch measurement theory. *Psychiatr Rehabil J* 2015; **38**: 349–58.

