Non-parametric measurement for patient-reported outcomes

Boehnke, Jan R.

Published in:
Quality of Life Research

DOI:
10.1007/s11136-021-03073-z

Publication date:
2022

Document Version
Peer reviewed version

Editorial commentary on the Special Section

"Non-parametric Measurement for Patient Reported Outcomes"

Jan R. Boehnke[1]

[1] School of Health Sciences, University of Dundee, Dundee, UK

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in "Quality of Life Research", and is available online at https://doi.org/10.1007/s11136-021-03073-z.

Corresponding author:
Jan R. Boehnke
School of Health Sciences
City campus
11 Airlie Place
University of Dundee
Dundee
DD1 4HJ
United Kingdom
j.r.boehnke@dundee.ac.uk

We issued the call for the set of papers published in this special section in 2019 with the goal to illuminate innovative state-of-the-art methods for nonparametric measurement. The original idea for this call goes back to a series of potential topics for methodological special sections in our journal collated by our previous Co-Editor in Chief, Frans J. Oort, which we discussed and developed further in our editorial meetings at the International Society for Quality of Life Research (ISOQOL) Annual Conferences in Philadelphia (2017) and Dublin (2018). And I am very grateful to our two excellent guest editors, Klaas Sijtsma and Andries van der Ark, who helped us to finalize the call and who facilitated the project. We received 13 expressions of interest and six full papers were submitted, of which five are published in this special section.

While parametric item response models are widely used (for example more than 7,007 search results in PubMed titles and abstracts at the date of writing; 264 in Quality of Life Research), a search for terms specifically related to non-parametric item response models reveals only 392 results (15 in Quality of Life Research, including the papers of this issue). Nonparametric item response models have at least been around as long as the parametric model family as for example seen in Guttman's work on scalogram analysis [1]. But owing to their versatility and solutions for practical problems, parametric models were the main focus of research and application until a renewed interest in the 1980ies [2]. This may explain some of the differences in uptake, but originally limited options for the modelling of polytomous responses [3], delayed development of a minimal set of standard analyses and that the approaches may have appeared less "canonized" and were less accessible than their parametric alternatives [4,5] may have contributed as well. Today, nonparametric models offer a theoretical framework to study human response processes, to explore the underlying assumptions of item response models, and procedures to explore data quality [2,6].

While the last years have seen a swiftly increasing popularity of nonparametric item response theory (NP-IRT) to investigate health-related quality of life (HRQL) data, applications remain limited and the full potential of these models has not been explored for our field. The introductory paper by the guest editors [7] starts with a mention of previous applications and moves on to provide an overview of connections to parametric models and a series of steps required for applying NP-IRT. The steps are illustrated with a worked example and the text provides references to more detailed introductions and tutorials. The authors illustrate well that NP-IRT models are flexible and that this could be a distinct advantage when modeling assessments of HRQL and patient-reported outcome data, where our theories may not yet have matured enough to justify the choice of particular parametric models.

The section starts with the application of Mokken scale analysis to the EQ-5D-5L using data from people living with and without chronic conditions in six Western countries [8]. The paper firstly demonstrates the use of the methods that are the focus of the special section with data relevant to our field. But it secondly also asks an important question about whether there is potential in analyzing a

measure that derives its validity mainly from valuation studies estimating preference-based scoring weights with methods that could in contrast potentially support the use of a summed score for further analysis. The result is favorable in this application, but with respect to the anxiety/depression dimension the results underscore the importance of reference samples when evaluating psychometric properties. The team suggests that this item might need to be evaluated separately. The authors discuss results and approaches in detail, and they highlight that the ultimate decision about scoring methods needs to be based in a theoretical framework: Positive results obtained with a psychometric procedure do not provide an answer what a score "means". The paper is accompanied with a detailed set of R syntaxes.

A second paper illustrates how applications of parametric item response models can benefit from parallel use of multiple models to explore item response functions. Based on PROMIS Wave 1 Physical Functioning data, the paper compares results obtained with the parametric Graded Response Model with those from a model employing monotonic polynomials [9]. The paper illustrates well that monotonic polynomials can be used to detect item misfit and that this additional information can be used for item bank and computerized adaptive test development. The argument is supported by a simulation study investigating recovery accuracy of several model parameters and practical implications are discussed. The extensive online materials accompanying the paper cover additional explanations, results, and R syntax for the reported analyses.

Two papers offer important extensions and clarifications of NP-IRT especially for the HRQL context. Data in our field are often collected in contexts that can violate the assumption that observations are independent of each other (clustering or nesting introduced by e.g., health services, schools, geographic regions). In this case, a range of the statistics relevant for the psychometric analyses are inappropriate and the first of these papers offers a solution within the context of automated item selection in Mokken scale analysis [10]. Accompanied with in-text R syntax examples and a detailed set of materials (https://osf.io/y7xud/), the paper firstly offers point estimates and standard errors for scalability coefficients for clustered data; and it secondly extends the automated item selection procedure with a formal test of whether a minimal scalability coefficient is exceeded (taking uncertainty of the point estimate into account). The second paper [11] explores the *Crit* coefficient, which is an index of violations of common assumptions in Mokken scale analysis. Discussed for dichotomous items, the two simulation studies and worked example show that the index performs well for certain conditions. An incidental result of the study is that robust Mokken scale analysis requires substantial sample sizes [12]. Although likely for more advanced users, the applied example and discussion of practical suggestions provide more widely applicable insights into how to conduct Mokken scale analysis.

The final paper of the special section [13] extends a line of studies [14] investigating the use of item response models to explore response shifts [15,16]. The team introduces a model including a set of assumptions under which recalibration response shift should be reflected in the number of "Guttman

errors", essentially the number of violations of item ordering from the easiest to the most difficult item of a scale observed in a respondent's response vector. The procedure is explored in a simulation study as well as an illustrative example, demonstrating its potential usefulness in certain situations (including detailed materials and Stata syntax: https://osf.io/h9nyd/). But more importantly, the presentation is an excellent vehicle to discuss the connections between different models, response shift detection methods, and their expected results, illustrating the potential of multi-model perspectives to generate insights into the response shift phenomenon.

Overall, the papers offer a range of insights into NP-IRT models, different theoretical and applied perspectives, as well as a range of materials to start working with these models. Item response models are a key psychometric tool for both development/ evaluation purposes as well as for deployment of instruments in HRQL research and practice. As the papers show, extending our toolbox with NP-IRT models offers potential for insights that can add to our understanding of how people use questionnaires to communicate about HRQL as well as shed new light on how substantive HRQL theories connect with the data we gather.

REFERENCES

1. Guttman, L. (1947). The Cornell technique for scale and intensity analysis. *Educ Psychol Meas, 7*(2), 247-279, doi:10.1177/001316444700700204.
2. van der Linden, W. J., & Hambleton, R. K. (1997). Nonparametric models. In *Handbook of modern item response theory* (pp. 347-349). New York: Springer.
3. Sijtsma, K., & Van der Ark, L. A. (2001). Progress in NIRT analysis of polytomous item scores: Dilemmas and practical solutions. In A. Boomsma, M. A. J. Van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 297-318). New York: Springer.
4. Sijtsma, K., & van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *Br J Math Stat Psychol, 70*(1), 137-158, doi:10.1111/bmsp.12078.
5. Junker, B. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M. A. J. Van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 247-276). New York: Springer.
6. Meijer, R. R., Tendeiro, J. N., & Wanders, R. B. K. (2015). The use of nonparametric item response theory to explore data quality. In *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 85-110). New York: Taylor & Francis (Routledge).
7. Sijtsma, K., & van der Ark, L. A. (2021). Advances in nonparametric item response theory for scale construction in quality-of-life research. *Qual Life Res*, doi:10.1007/s11136-021-03022-w.
8. Feng, Y. S., Jiang, R., Pickard, A. S., & Kohlmann, T. (2021). Combining EQ-5D-5L items into a level summary score: demonstrating feasibility using non-parametric item response theory using an international dataset. *Qual Life Res*, doi:10.1007/s11136-021-02922-1.
9. Falk, C. F., & Fischer, F. (2021). More flexible response functions for the PROMIS physical functioning item bank by application of a monotonic polynomial approach. *Qual Life Res*, doi:10.1007/s11136-021-02873-7.
10. Koopman, L., Zijlstra, B. J. H., & van der Ark, L. A. (2021). A two-step, test-guided Mokken scale analysis, for nonclustered and clustered data. *Qual Life Res*, doi:10.1007/s11136-021-02840-2.
11. Crisan, D. R., Tendeiro, J. N., & Meijer, R. R. (2021). The Crit coefficient in Mokken scale analysis: a simulation study and an application in quality-of-life research. *Qual Life Res*, doi:10.1007/s11136-021-02924-z.
12. Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum Sample Size Requirements for Mokken Scale Analysis. *Educational and Psychological Measurement, 74*(5), 809-822, doi:10.1177/0013164414529793.
13. Dubuy, Y., Sebille, V., Grall-Bronnec, M., Challet-Bouju, G., Blanchin, M., & Hardouin, J. B. (2021). Evaluation of the link between the Guttman errors and response shift at the individual level. *Qual Life Res*, doi:10.1007/s11136-021-03015-9.
14. Blanchin, M., Sebille, V., Guilleux, A., & Hardouin, J. B. (2016). The Guttman errors as a tool for response shift detection at subgroup and item levels. *Qual Life Res, 25*(6), 1385-1393, doi:10.1007/s11136-016-1268-8.
15. Rapkin, B. D., & Schwartz, C. E. (2019). Advancing quality-of-life research by deepening our understanding of response shift: a unifying theory of appraisal. *Quality of Life Research*, doi:10.1007/s11136-019-02248-z.
16. Vanier, A., Oort, F. J., McClimans, L., Ow, N., Gulek, B. G., Bohnke, J. R., et al. (2021). Response shift in patient-reported outcomes: definition, theory, and a revised model. *Qual Life Res, 30*(12), 3309-3322, doi:10.1007/s11136-021-02846-w.