

Letter to the Editor

Why psychometrics is important - A Response To: Aubeeluck, Buchanan & Stuppel (2013) Journal of Huntington's Disease 2(4) 453-454

Peter Hagell^a and Steve Smith^b

^a*The PRO-CARE Group, School of Health and Society, Kristianstad University, Kristianstad, Sweden*

^b*School of Nursing Sciences, Faculty of Medicine and Health, University of East Anglia, Norwich, England, UK*

Dear Editor,

We thank Aubeeluck, Buchanan and Stuppel (ABS) for their comments [23] regarding our recent paper on the psychometric properties of the Huntington's Disease Quality of Life Battery for Carers (HDQoL-C) and the Alzheimer's Carers Quality of Life Inventory (ACQLI) among carers of people with Huntington's disease (HD) [1], and the Editor for providing us with this opportunity to clarify the issues raised in the letter.

It is true that our sample size was small. This was acknowledged [1], together with the fact that studies [2] have shown that samples of this size (or smaller; $n \geq 20$) can be useful and provide stable interpretations relative to larger samples for basic traditional psychometric analyses. For example, the FACIT-F, a widely used fatigue scale, was initially tested with a sample of $n = 49$ [3]; subsequent larger-scale studies have not contradicted initial results.

Arguing that single-centre samples are problematic means that single-centre studies are inappropriate for the HDQoL-C. This is because traditional psychometric properties are not fixed scale characteristics but sample dependent, and should be supported within the sample in which a scale is used [4–6]. For example, the developers of the SF-36, one of the most widely used rating scales in the health sciences, argue that

basic scaling assumptions should always be met before interpreting SF-36 data [7, 8].

We find the comment that our data are homogenous curious. First, and relevant also to the previous point, our sample characteristics were very similar to those previously reported in psychometric HDQoL-C studies [9, 10]. Secondly, all HDQoL-C scores ranged across their full or close to full possible ranges [1]. This aspect is important, since failure to do so limits the interpretability of psychometric data due to lack of sample representation across the latent continuum [11]. Had our HDQoL-C data been restricted to a narrow range, there would have been a problem related to “homogenous data”. This was not the case. Although our sample was recruited from a single centre, the recruitment area covered four English counties (a fact that unfortunately was unacknowledged in our paper).

We agree that factor analysis (FA) is common in psychometric studies. However, there are good arguments against factor analysing item-level data [12]. The prime reason for conducting item-level FA is to assess dimensionality [13]. Due to our limited sample size and inherent problems with such analyses, we took an alternative approach, multitrait/multi-item scaling analysis [11, 13]. Although addressing the same fundamental issue as FA, multitrait/multi-item scaling

has fewer assumptions and sources of misinterpretations. Our results suggested dimensionality problems with all HDQoL-C domain and total scores except the *Satisfaction with Life* domain. As pointed out [1], these results are in general agreement with previous HDQoL-C studies [9, 10]. In those studies, no analyses of the dimensionality of the full HDQoL-C were presented but principal component analyses (a form of FA) were applied to each proposed HDQoL-C domain individually. This yielded 2-3 components per domain, suggesting multidimensionality. While noted [9, 10], this has been unacknowledged in the recommended use of the HDQoL-C; instead, total domain scores and a total HDQoL-C score are recommended [14, 15]. This is in contrast to basic psychometric assumptions [12].

It is unclear what ABS mean by “standard tests of validity”. In essence, there are three related steps to traditional rating scale validation: (i) Defining and operationalizing the target variable(s); (ii) testing scaling assumptions and internal validity; (iii) testing *a priori* hypotheses regarding relationships between scores and external criteria (external construct validity). We considered the first of these and focused on the second [1]. This involved tests addressing whether items *within* proposed scales appear to represent an underlying latent variable, as well as tests of the legitimacy of calculating summed total scores according to traditional (Likert) psychometric theory [7, 16]. Such support is essentially a prerequisite for confident use and interpretation of scores [4–8, 11]. Assuming that “standard tests of validity” refers to external construct validity, we did not address this due to general lack of supported scaling assumptions/internal construct validity for the HDQoL-C, and because such tests provide generally weak and circumstantial evidence [17, 18].

This leads us to the conceptual underpinnings of the HDQoL-C. ABS argue that we are incorrect in saying that the HDQoL-C lacks explicit definitions of its target variables and that content validity was not informed by carers. However, nowhere in the HDQoL-C publications known to us are *Practical aspects of Caregiving*, *Satisfaction with Life* or *Feelings about Living with HD* explicitly defined. As these are the proposed HDQoL-C scores [14, 15], these are the target variables. As for the total HDQoL-C score, the target variable is supposedly quality of life (QoL), which ABS imply is based on the WHO definition of QoL. This is puzzling, since it is not mentioned in previous publications [9, 10, 19]; instead, the guiding conceptual framework referred to is that of Cummins [20]. The literature does state that the content of the HDQoL-C is

based on qualitative work involving HD carers [19] and that the three HDQoL-C domains were identified as pertinent to the QoL of spousal HD carers [9]. Indeed, HD carers were involved in the development of the HDQoL-C by (i) reviewing a generic QoL instrument (based on Cummins’ QoL definition) for relevance; (ii) documenting what compromises or enhances their QoL (as defined by Cummins); (iii) describing how caring affects their QoL; and (iv) commenting on the appropriateness, difficulty and clarity of HDQoL-C items [9, 19]. As for content validity of the HDQoL-C, “two experts in the field of QoL and two experts in the field of HD were asked to comment on the item content” [9].

Most psychometric methods, including those employed in our [1] and other HDQoL-C studies [9, 10] assume that items are observable manifestations of the latent variable. That is, variations in item responses should reflect variations in, e.g. QoL – not the other way around [13]. Failure to meet basic psychometric criteria and to provide evidence that scale generated numbers represent measures rather than just numerals suggests that the scale does not qualify as a *measurement* tool [18, 21]. The development and psychometric testing of the HDQoL-C [1, 9, 10, 19] provide reasons to question its appropriateness as a tool for *measurement*. This is concerning, particularly since it has been suggested as an outcome measure [9, 10].

While carers may welcome reflecting on aspects covered by the questionnaire, item response rates and questionnaire completion time provide reasons to consider if the HDQoL-C could be improved from a respondent perspective. HDQoL-C completion time in our study was very similar to that previously reported [9], and was >4 times longer than that of our comparator questionnaire [1]. As for response rates, missing item responses ranged from 0-11.5% [1]. It is unknown how this compares to previous studies since response rates were not reported [9, 10], although Cox (2002) reported a 17.5% overall HDQoL-C response rate [22].

In our study [1], we also tested a carer QoL questionnaire developed for carers of people with Alzheimer’s disease (the ACQLI) to further explore the notion that carer impact has similarities across neurodegenerative disorders. Given that the HDQoL-C failed to outperform the ACQLI, we concluded that there is “tentative support for the idea of a common cross-diagnostic carer impact of neurodegenerative disorders” and that “the ACQLI appears a promising candidate tool for further work towards such cross-diagnostic QoL assessments” [1]. This does not argue against disease specific instruments.

ABS criticise the fact that the ACQLI is not freely available on the Internet. This is true and common practice for reasons including quality assurance through protection against questionnaire modifications leading to different (unsupported/untested) versions and translations. Sample ACQLI pages are available on the Internet and the full questionnaire can be obtained for non-commercial use following signed agreement and a nominal administrative fee (www.galen-research.com).

We have accepted the invitation from ABS to share our data, provided that data are subjected to rigorous state-of-the-art analyses using modern psychometric methods (Rasch analysis), and offered to assist in conducting and interpreting such analyses. We hope to receive a positive response. We also hope to see further independent psychometric analyses of the HDQoL-C, since scale validation is an iterative quantitative and qualitative process involving theoretical, substantive and empirical considerations [4, 18]. If we take our carers, patients and studies seriously, we also need to be serious about the outcome measures we use. Unless rating scales are treated with full scientific rigour, advances in the clinical sciences will be hampered and opportunities to improve care may be lost.

REFERENCES

- [1] Hagell P, Smith S. A psychometric comparison of two carer quality of life questionnaires in Huntington's disease: Implications for neurodegenerative disorders. *J Huntingtons Dis*. 2013;2(3):315-22.
- [2] Hobart JC, Cano SJ, Warner TT, Thompson AJ. What sample sizes for reliability and validity studies in neurology? *J Neurol*. 2012;259(12):2681-94.
- [3] Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *J Pain Symptom Manage*. 1997;13(2):63-74.
- [4] Food and Drug Administration. Patient-Reported Outcome measures: Use in Medicinal Product Development to Support Labelling Claims. Washington DC, 2009.
- [5] Hagell P, Tornqvist AL, Hobart J. Testing the SF-36 in Parkinson's disease. Implications for reporting rating scale data. *J Neurol*. 2008;255(2):246-54.
- [6] Thompson B, Vacha-Haase T. Psychometrics is datametrics: The test is not reliable. *Educ Psychol Meas*. 2000;60(2):174-95.
- [7] Ware JE Jr., Harris WJ, Gandek B, Rogers BW, Reese PR. MAP-R for Windows: Multitrait/multi-item analysis program - revised user's guide. Health Assessment Laboratory, Boston, 1997.
- [8] Ware JE, Jr., Kosinski M, Gandek B. SF-36 health survey: Manual and interpretation guide. Lincoln, RI: QualityMetric Incorporated. 1993;2000.
- [9] Aubeeluck A, Buchanan H. The Huntington's disease quality of life battery for carers: Reliability and validity. *Clin Genet*. 2007;71(5):434-45.
- [10] Aubeeluck A, Dorey J, Squitieri F, Clay E, Stuppel EJ, De Nicola A, et al. Further evidence of reliability and validity of the Huntington's disease quality of life battery for carers: Italian and French translations. *Qual Life Res*. 2013;22(5):1093-8.
- [11] 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):iii, ix-x, 1-177.
- [12] Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. McGraw-Hill, Inc. New York, 1994.
- [13] Fayers PM, Machin D. Quality of life: The assessment, analysis and interpretation of patient-reported outcomes. Second edition. John Wiley & Sons, Ltd., West Sussex, 2007.
- [14] Aubeeluck A, Buchanan H. Huntington's Disease Quality of Life Battery for Carers (HDQoL-C) - Validated for use with family Carers of Persons with Huntington's Disease [user guide]. University of Nottingham, Nottingham, UK, 2007.
- [15] Aubeeluck A, Dorey J, M. T, Buchanan H. Huntington's Disease Quality of Life Battery for Carers - Short Form (HDQoLC-SF) - Validated for use with Family Carers of Persons with Huntington's Disease. User guide for English and other translations. University of Nottingham, Nottingham, UK, 2009.
- [16] Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;140:1-55.
- [17] Hobart J, Cano S, Baron R, Thompson A, Schwid S, Zajicek J, et al. Achieving valid patient-reported outcomes measurement: A lesson from fatigue in multiple sclerosis. *Mult Scler*. 2013;19(13):1773-83.
- [18] Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *Lancet Neurol*. 2007;6(12):1094-105.
- [19] Aubeeluck A, Buchanan H. A measure to assess the impact of Huntington's disease on the quality of life of spousal carers. *British Journal of Neuroscience Nursing*. 2006;2(2):88-95.
- [20] Cummins RA. The Comprehensive Quality of Life scale - Adult, Fifth edition (ComQoL-A5) Manual. Deakin University, Melbourne, 1997.
- [21] Stenner AJ, Stone MH, Burdick DS. Indexing vs. measuring. *Rasch Meas Trans*. 2009;22(4):1176-7.
- [22] Cox M. Quality of life among carers of people with Huntington's disease. *British Journal of Neuroscience Nursing*. 2012;8(5):288-94.
- [23] Aubeeluck A, Buchanan H, Stuppel E. A psychometric comparison of two carer quality of life questionnaires in Huntington's disease: Implications for neurodegenerative disorders - A response to: Hagell and Smith (2013). *Journal of Huntington's Disease*. 2013;2(4):453-4.