



Open

## Response to Rubanovich et al.

Rubanovich et al.<sup>1</sup> raise several concerns regarding the development of the Clinician-reported Genetic testing Utility InDEx (C-GUIDE).<sup>2</sup> We appreciate the opportunity to deepen the discussion of measurement science as it applies to genomic medicine. The authors assert that we “did not align our approach with psychometric best practices” and that we “did not describe or refer to any theoretically sound methodology for scale development.” Although not stated explicitly in the paper, we thoroughly debated the assumptions and approaches associated with both the traditional psychometric and the more relevant and clinically oriented clinimetric approach to measurement.<sup>3</sup> While the well-established toolkit provided by the psychometric approach is appealing, the purpose and target population for which C-GUIDE is intended precisely align with the clinimetric approach to measurement development.

The aim of a psychometric approach is to create a scale in which multiple items measure a construct of interest that exists, but is not directly measurable. This is called a latent variable, the structure of which is determined by the statistical relationship among the items that reflect it. A large number of items are generated and statistics are used to establish a set of highly correlated items, representing a reflective model.<sup>4–6</sup> In contrast, a clinimetric approach relies on notions of clinical relevance and clinical sensibility;<sup>3</sup> clinimetric tools are not constrained to reflective items or statistical relationships among them. Rather, they include causal or formative items. In a formative model, the concept being measured does not exist; rather, it is constructed through a process of identifying items that form it. Clinical utility is a clinimetric construct that has been defined (i.e., formed) by experts in the field, beginning with the ACCE model which assesses the Analytic validity, Clinical validity, Clinical utility, and Ethical, legal and social implications of genetic tests.<sup>7</sup> Outcomes related to diagnosis, management, health benefits, and family testing, for example, constitute this construct. We, as a community of practice, have formed this construct; the C-GUIDE aims to provide a clinically sensible way of quantifying it.

In essence, psychometric and clinimetric approaches represent different theoretical perspectives, aims, and measurement development pathways. Often, different measurement tools emerge. For example, Juniper<sup>8</sup> developed an asthma-related quality of life tool and contrasted psychometric methods using factor analysis with a clinimetric approach using patients’ opinions of importance. Psychometric analysis resulted in a 36-item instrument

compared with 32 items by the “importance” method. Only 20 items were common to both. The “importance” method was chosen for the final instrument as it showed better correspondence to clinical sensibility. The psychometric method would have excluded three items of greatest importance to the stakeholders and other important items relating to functional impairment.

We concur that a nomological network is appropriate for establishing the structure of a psychometric measure; the same requirement, however, does not apply to clinimetrics. The nomological net assumes the construct “exists” and the structure, as derived by the network, establishes the laws for what “it” is.<sup>9</sup> A formative construct does not truly exist; we formulate it to meet a test of clinical sensibility. Further, Rubanovich et al. suggest that we should have generated many more items in the early stages of our work than we reported. In fact, the 25 preliminary items were culled from a longer list. Each item was worded several different ways and iteratively modified, where applicable, to suit the context of secondary and pharmacogenomics variants. As such, our preliminary list in all of its permutations far exceeded 25. Since the goal of a formative model, however, is to identify distinct aspects that form the construct, an extensive list of interchangeable items is not required in the way it is for a reflective measure. The authors also suggest that we “changed the model of our construct in response to feedback on specific items.” The “structure” of C-GUIDE was not derived *de novo*, but was informed by the Fryback and Thornbury<sup>10</sup> model, relevant literature, and qualitative interviews. While the items constitute the construct, they do not form formally structured domains. The consolidation of the family and psychosocial impact “domains” did not relate to poorly written items; rather it related to respondents’ sentiment that familial implications could influence and be influenced by psychosocial impact. As such, these domains were not distinct. In the end, while C-GUIDE attends to diagnostic, management, and familial/psychosocial aspects of clinical utility, these domains do not represent the formal internal structure of a construct in a psychometric sense. They represent literature-derived, clinically sensible components of clinical utility, finer-grained aspects of which are represented by the index items.

The decision to implement a total sum rather than domain scores was precisely because some domains—as Rubanovich et al. point out—would be disproportionately weighted over others, with no evidence to justify this. With respect to establishing item weights, factor analysis is not an appropriate tool because correlation among items is not expected. As such, item weights are typically decided by respondents and researchers, and rely on clinical sensibility.<sup>6</sup> We presented our expert panel with three different weighting strategies. The

majority selected the strategy that assigned a weighted value to the items identified to be the most important. However, experts rated all items as highly important. As confidence intervals were overlapping, an item-weighted scoring system could not be justified. Also in keeping with a formative model, response options were informed by experts and were not statistically derived. Indeed, interval scales with 5–7 response options provide “more” information and are often viewed to be “statistically useful”—but only when differences between response options are clinically meaningful. While more response options may contribute to discriminative validity, they reduce evaluative validity. When developing a clinimetric tool, the type and number of response options depends on clinical relevance and respondents’ ability to meaningfully decipher the options. While we included a 5-point Likert scale scoring option, expert feedback indicated that many items could not be deciphered at this level of granularity and that simple categorical or ordinal response options would be easier to complete and interpret.<sup>6</sup>

Rubanovich *et al.* critiqued our inclusion of clinicians’ perceptions of psychosocial impact, claiming their perception may not accurately reflect families’ perceptions. Our clinician experts advocated for the inclusion of these items, considering them highly relevant to the construct. We did not suggest that C-GUIDE could be used in lieu of familial or individual measures of perceived utility. Indeed, patient-reported measures of utility are of utmost importance to supplement and complement clinician perceptions. Finally, it is unfortunate that Rubanovich *et al.* interpreted our developmental work as ready for prime time. Indeed C-GUIDE version 1.0, named as such for its preliminary nature, is currently undergoing further field testing, validity, and reliability testing. Version 1.0 is not intended as an outcome measure until evidence of its performance justifies doing so. We appreciate the interest of Rubanovich *et al.* in our work and for expanding the discussion of the measurement of clinical utility among the clinical and research community.

## ACKNOWLEDGEMENTS

This work was funded by the Canadian Institutes of Health Research (CIHR).

## DISCLOSURE

The authors declare no conflicts of interest.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

R. Z. Hayeems, ScM, PhD<sup>1,2</sup>, S. Luca, MA<sup>1</sup>,  
W. J. Ungar, MSc, PhD<sup>1,2</sup>, A. Bhatt, BMSc(Cand)<sup>1</sup>,  
L. Chad, MDCM, MHSc<sup>3,4</sup>, E. Pullenayegum, PhD<sup>1,5</sup> and  
M. S. Meyn, MD, PhD<sup>6</sup>

<sup>1</sup>Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada; <sup>2</sup>Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada; <sup>3</sup>Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; <sup>4</sup>Department of Pediatrics, University of Toronto, Toronto, ON, Canada; <sup>5</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; <sup>6</sup>Center for Human Genomics and Precision Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA. Correspondence: R. Z. Hayeems (robin.hayeems@sickkids.ca)

## REFERENCES

- Rubanovich CK, Stevens Smith H, Bloss CS. Response to Hayeems *et al.* 2019: The development of the Clinician-reported Genetic testing Utility InDEx (C-GUIDE). *Genet Med.* (in press).
- Hayeems RZ, Luca S, Ungar WJ, *et al.* The development of the Clinician-reported Genetic testing Utility Index (C-GUIDE): a novel strategy for measuring the clinical utility of genetic testing. *Genet Med.* 2019 Jul 31; 10.1038/s41436-019-0620-0 [Epub ahead of print].
- Feinstein AR. *Clinimetrics.* New Haven, CT: Yale University Press; 1987.
- Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. *J R Stat Soc Ser A.* 2002;165:233–61.
- Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. *Psychol Methods.* 2000;5:155–174.
- de Vet HCW, Terwee CB, Mokkink LB, *et al.* *Measurement in medicine: a practical guide.* Cambridge, UK: Cambridge University Press; 2011.
- Haddow JEPG. A model process for evaluating data on emerging genetic tests. In: Khoury MLJ, Burke W, editors. *Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease.* Oxford, UK: Oxford University Press; 2003. p. 217–233.
- Juniper EF, Guyatt GH, Streiner DL, *et al.* Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol.* 1997;50:233–238.
- Cronbach LJP. Construct validity in psychological tests. *Psychol Bull.* 1955;52:281–302.
- Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making.* 1991;11:88–94.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2019

Advance online publication 19 November 2019. doi:10.1038/s41436-019-0700-1