



Commentary on the development of the Clinician-reported Genetic testing Utility InDEx (C-GUIDE)

To the Editor:

Hayeems and colleagues recently published an article on the development of a novel measure of clinical utility titled the Clinician-reported Genetic testing Utility InDEx (C-GUIDE)^{TM(1)}. C-GUIDE is an 18-item tool with 9 supplemental questions for clinical context. The authors note that the field currently lacks a “single validated measure that quantifies clinical utility”¹ and that the C-GUIDE was developed to address this gap. We commend the authors in their efforts to develop a tool to measure clinicians’ perceptions of clinical utility in genomics given the current lack of validated measures in the field coupled with the importance of assessing utility to evaluate the use of genomics in clinical practice. As researchers conducting instrument development in our own work, we read the article with great interest. Unfortunately, the C-GUIDE seems to be an underdeveloped measure, the design of which neglected many psychometric best practices. We are concerned that C-GUIDE’s dissemination and wide use in its current form could adversely impact the field, particularly given that little evidence of its reliability or validity was presented.

A major issue with the C-GUIDE is that it appears that the authors have conflated defining a conceptual model with developing a measure. Scale development best practices call for first defining the construct one is trying to measure, including the development of a nomological net,² and then conducting item development.^{3–6} Establishing a nomological net, which includes the internal structure of the construct (including subconstructs) and a description of the construct’s relationship to external constructs, helps ensure the development of a reliable and valid measure of the construct of interest and low measurement error.^{3,4} In contrast, Hayeems et al. change the model of their construct in response to feedback on specific items. For example, the authors describe consolidating two subconstructs (i.e., family implications and psychosocial implications) based on respondent answers to individual items. It would be more appropriate to rewrite and add new items to ensure adequate representation of the subconstructs that comprise the model, which helps ensure structural and content validity, rather than reconceptualizing the model of the construct^{3,4} based on poorly written items. This is particularly true given that their literature review and stakeholder interviews indicated five separate domains or subconstructs of clinical utility.

There are other methodological issues of concern in the C-GUIDE development related to the number of items that were developed as well as the scoring procedures for the instrument. Specifically, instrument development best practices call for initially writing many more items than one plans to include in the final measure (i.e., 2–5 times as many), which would have allowed for a more robust item elimination process^{3,4} in this case. Moreover, while an instrument that produces a single score to reflect overall impact or utility is theoretically appealing and would be a substantial advancement over the multicategory checklists used in previous studies,⁷ the scoring procedure used to arrive at the C-GUIDE’s single total score is problematic because the item count per subconstruct varies. Therefore, the C-GUIDE’s single total score disproportionately weights one subconstruct over another and does not represent the subconstructs equally. In addition, it appears that the authors asked respondents to choose a scoring approach without themselves first taking quantitative considerations into account. For instance, over 70% of the tool’s items are two- or three-point response options, which would lead to lower reliability than four-, five-, six- or seven-point scales⁸ and could limit the tool’s utility in statistical analyses.

We agree with the authors that the potential impact on family members and nonclinical domains is important to include in any score that attempts to summarize the impact of genomic sequencing, but it is noteworthy that per the C-GUIDE’s current administration, familial and psychosocial implications are assessed by the clinician.⁹ However, because clinicians’ perceptions of family impact may not accurately reflect families’ perceptions, the instrument should not be used in lieu of familial or psychosocial measures of perceived utility without first showing that clinician perceptions as assessed via the C-GUIDE are correlated with familial perceptions (i.e., criterion validity).

Finally, it is not advisable to widely disseminate a measure without first showing compelling evidence of the measure’s reliability and validity (e.g., test–retest reliability, interrater reliability, concurrent and predictive criterion validity, convergent and divergent validity, structural validity, sensitivity to change).⁴ While it is theoretically possible for the measure, in its current form, to show compelling evidence of reliability and validity in future studies, we are concerned that the problems we have outlined in the measure development process may contribute to limited reliability and validity overall. Specifically, while the authors cited Fryback and Thornbury’s hierarchical model of efficacy¹⁰ as the basis for developing their construct of clinical utility, they did not describe or refer to any theoretically sound methodology for scale development. The authors indicate their intention to conduct future work in reliability and validity, however, little evidence is presented in the current publication even though

the tool's publication and dissemination give the impression that C-GUIDE is ready for use by others in the field.

There is a dearth of tools available to assess clinical utility, and thus, developing psychometrically sound measures of this construct is critical for evaluating the use of genomics in clinical practice. We commend the authors on their review of the literature and on gathering feedback from clinicians in their process. Indeed, incorporating findings from the literature as well as engaging experts and a measure's target population are among some of the recommendations from scale development best practices.^{3,4} Nevertheless, given the methodological deficiencies we have outlined here, C-GUIDE, in its current form, appears underdeveloped and its use by others in the field is likely premature.

ACKNOWLEDGEMENTS

The authors thank Vanessa Malcarne and Amy McGuire for their review and critique during the drafting stage of this correspondence.

FUNDING

This work was supported by two grants from the National Human Genome Research Institute (R01 HG008753, principal investigator: Bloss; and U01 HG006485-S1, principal investigator: McGuire).

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.

Caryn Kseniya Rubanovich, MS¹,
Hadley Stevens Smith, PhD, MPSA² and
Cinnamon S. Bloss, PhD³

¹Joint Doctoral Program in Clinical Psychology, San Diego State University/University of California–San Diego, San Diego, CA, USA; ²Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX, USA; ³Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. Correspondence: Cinnamon S. Bloss (cbloss@ucsd.edu)

REFERENCES

1. Heyeems R, Luca S, Ungar W, 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 [Epub ahead of print].
2. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52:281.
3. Boateng GO, Neilands TB, Frongillo EA, et al. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health*. 2018;6:149.
4. DeVellis RF. *Scale development: Theory and applications*. Vol 26: University of North Carolina, Chapel Hill, NC: Sage publications; 2016.
5. Raykov T, Marcoulides GA. *Introduction to psychometric theory*. New York, NY and Hove, East Sussex: Routledge; 2011.
6. MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. *MIS Q*. 2011;35:293–334.
7. Scocchia A, Wigby KM, Masser-Frye D, et al. Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphology clinic in Mexico. *NPJ Genom Med*. 2019;4:5.
8. *Survey Measurement and Process Quality*, Lyberg, L, Biemer, P., Collins, M., deLeeuw, E., Dippo, C., Schwarz, N., Trewin, D. (eds.), New York: John Wiley; 1997:141–64 <https://doi.org/10.1002/9781118490013>
9. Smith HS, Russell HV, Lee BH, Morain SR. Using the Delphi method to identify clinicians' perceived importance of pediatric exome sequencing results. *Genet Med*. 2019 Jul 5 [Epub ahead of print].
10. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11:88–94.

Advance online publication 19 November 2019. doi:10.1038/s41436-019-0701-0