Why a global PROMIS® can't be kept

Jack DOWIE [ab1] and Mette Kjer KALTOFT [b]

[a] *London School of Hygiene and Tropical Medicine*
[b] *University of Southern Denmark*

Abstract. Composite multi-dimensional constructs, such as 'global mental health' and 'global physical health', in PROMIS® instruments and ICHOM standard outcome sets, are formative, not reflective. Their preference-insensitivity means they are potentially misleading in both clinical and policy decision making practice. Their frequent validation by reflective psychometric tests is also improper methodologically The spread of these instruments is occurring without sufficient awareness on the part of patients, clinicians, researchers and policy makers that the need for group-specific preference bases ('tariffs') for such measures rules out any possibility of 'international gold standard metrics'.

Keywords: Patient Reported Outcome Measures; preference-sensitive, formative, reflective, PROMIS

Introduction

"PROMIS® is increasingly recognized as the international gold standard for patient-centered assessment…" suggest Evans and colleagues [1] (p345). PROMIS® is a large and expanding bank of patient-related outcome items and instruments, operating under the aegis of the PROMIS Health Organisation (PHO), with an associated user community. PROMIS® is the registered trademark of the U.S. Department of Health and Human Services. (http://www.promishealth.com).

"The Patient-Reported Outcomes Measurement Information System (PROMIS®) is a National Institutes of Health initiative to develop state-of-the-science self-report measures to assess functioning and well-being in physical, mental, and social domains of health. PROMIS measures are potentially useful to screen for disability, identify health care disparities, enhance communication between patients and clinicians, and improve population health… PROMIS includes item banks that can be administered using computer-adaptive testing, short forms for individual domains, and profiles that yield information about multiple domains for use in clinical trials, observational studies, and clinical practice. The PROMIS-29 v2.0 profile measure assesses pain intensity using a single 0–10 numeric rating item and seven health domains (physical function, fatigue, pain interference, depressive symptoms, anxiety, ability to participate in social roles and

[1] Corresponding author: Jack Dowie, LSHTM, London WC1H 9SK; E-mail: jack.dowie@lshtm.ac.uk

activities, and sleep disturbance) using four items for each domain. [It] is analogous to the most widely used profile measure to date, the SF-36. But the PROMIS-29 v2.0 profile items were selected from PROMIS item banks calibrated using item response theory (IRT) analyses and all items in a domain are scored on the same underlying metric… While profile measures yield a wealth of information, higher-order summary measures [such as PROMIS GH 29 and PROMIS GH 10] are also useful." [2] (p1885-6).

The goals of PHO are to 'provide reliable, valid, and cost-effective measurement of relevant health outcomes to the greater scientific and clinical research community and to other health care organisations', and to increase their 'clinical adoption by organizing and presenting PRO data that are relevant and useful to clinicians, patients, and researchers'. Its aspirations include 'developing PROMIS into a gold-standard outcome metric' and making it 'part of routine clinical practice across multiple specialties'.

The subsequent adoption of PROMIS® global measures by the International Consortium for Health Outcomes Measurement (http://www.ichom.org), can be seen as a response to calls by ICHOM founder Michael Porter [3], to end the definition of quality in healthcare as compliance with evidence-based practice guidelines; re-defining it as improvement in patient-relevant outcomes and thereby ending the 'outcome-measurement paralysis' that has been the provider-based obstacle to 'value-based care': "… to unlock the potential of value-based health care for driving improvement, outcomes measurement must accelerate. That means committing to measuring a minimum sufficient set of outcomes for every major medical condition - with well-defined methods for their collection and risk adjustment - and then standardizing those sets nationally and globally." [3] (p504-5).

The aim of this paper is to alert those contemplating such development and/or use of a universal PROM (e.g. the adoption of a PROMIS® global measure) to its incompatibility with the acceptance of international (and intra-national) difference in preferences.

Method

Unfortunately the development of PROMs, has been, and still is, occurring without acknowledging that many of the latent (unobservable) constructs such as 'global mental health' and 'global physical health' in PROMIS® (and adopted in several ICHOM standard outcome sets) are *formative* not *reflective*. Apart from their validation as if they are reflective being improper methodologically, their preference-insensitivity potentially undermines decision making aimed at optimal person-centred care [4].

An inflamed appendix, e.g. appendicitis, is a reflective construct. It is 'reflected' in its various signs and symptoms - indicators, criteria, cues - such as rebound tenderness. Crucially the inflamed appendix can be operated on directly and this will, if successful, cause the signs and symptoms to disappear, along with construct (in this patient).

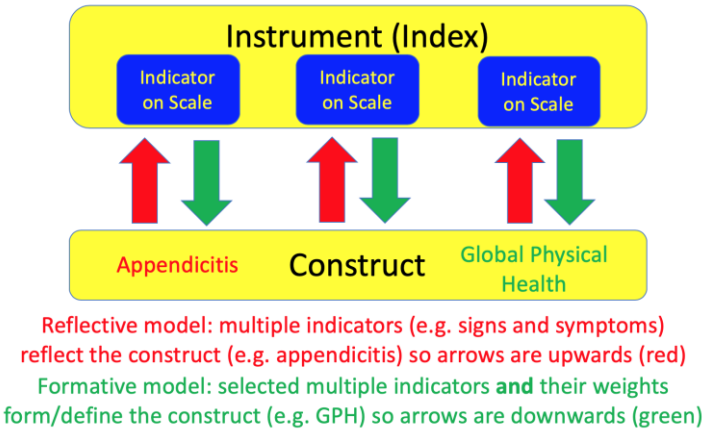

Figure 1. Comparison of formative and reflective constructs

'Global physical health', like most index measures in healthcare, including all multi-criterial PROMs, is a formative construct. It is 'formed' by the set of indicators selected to define it, such as 'everyday physical activity ability', 'pain' and 'fatigue' in the case of PROMIS GH 10. These three indicators need not be highly, or at all, inter-correlated, 'internal consistency' being a necessary condition only for a reflective construct. Moreover, removing/replacing one of them will change the construct, whereas removing/replacing any indicator of a reflective construct does not change it, any indicator being a reflection of the construct, not part of its definition. Crucially 'global physical health' cannot be affected directly, only by changing its indicators. (An online interactive version of PROMIS GH 10 is available at http://orthotoolkit.com/promis-10.)

So PROMIS GH 10 and 29 define and measure the multi-dimensional formative constructs of 'global physical health' that Promis GH 10 and 29 measure. Neither measures a thing called global physical health, which does not exist until it is constructed. They are measures of two different constructs, not two different measures of the same construct. If many agree to use them, given the embedded preferences, the source of any validity is dependent on that intersubjective agreement.

Generic instruments developed by economists for use in economic evaluation are the most prominent examples of acceptance that multi-dimensional indexes, such as Health-Related Quality of Life (HRQOL), are formative constructs. This is acknowledged in accepting that any HRQOL measure is preference-sensitive, so that a population-based 'tariff' is needed for policy use. Crucially *Danish HRQOL as measured by EQ-5D-5L using the Danish tariff* is a different construct from *French HRQOL as measured by EQ-5D-L using the French tariff*. They are *not* two measures of the same construct, i.e. HRQOL as defined by EQ-5D-5L). While one can compare the two measures, applying the very different French tariff in Denmark - or vice versa – doing so would only confirm the irrelevance of the comparison for any decision in either.

Result

One example is sufficient to confirm the issue with preference-insensitive 'global' PROMs. Using PROMIS GH 10 in their study of stroke patients, Katzan et al. noted that 'measures that make up the physical component score are less highly correlated with each other than the measures that make up the mental component score' [5] (p150). While 'the results of our study support the recommendation from ICHOM to use PROMIS GH as part of the standard set of outcome measures in stroke… [because] of the moderate internal reliability and poor model fit of summarizing items into physical and mental health component scores, greater focus should be placed on individual PROMIS GH items than on the component scores in patients with stroke.' [5] (p153). Given that formative indexes are weighted scales, their main problem lies in the lack of a credible, or any, preference basis, not such statistical considerations.

The case-mix origins of the 'internationalisation' problem is well-recognised, but not the preference-mix one. "The mean and standard deviation of all PROMIS scales are anchored on the US population… An open question for the future therefore remains: should we anchor scales based on the US general population, the respective country's population, or even on a global level?' [6] (p1010). Given these three options, only the country-specific one respects international heterogeneity in preferences.

Conclusion

PROMIS® and ICHOM fail to recognise that a 'gold standard' measure of any preference-sensitive formative construct is impossible. As with established generic measures of health-related quality of life, PROM researchers need to develop country-specific tariffs to ensure their measures reflect international differences in preferences.

References

[1] J.P. Evans, A. Smith, C. Gibbons, J. Alonso, J.M. Valderas, The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK, *Patient Related Outcome Measures* **9** (2018), 345–52.

[2] R.D. Hays, K.L. Spritzer, B.D. Schalet, D. Cella, PROMIS®-29 v2.0 profile physical and mental health summary scores, *Quality of Life Research* **27** (2018), 1885–91.

[3] M.E. Porter, S. Larsson, T.H. Lee, Standardizing Patient Outcomes Measurement, *New England Journal of Medicine* **374** (2016), 504–6.

[4] M.K. Kaltoft, J.B. Nielsen, J. Dowie, Formative preference-sensitive measures are needed in person-centred healthcare at both clinical and policy levels, *European Journal for Person Centered Healthcare*, **5** (2017), 495-500.

[5] I.L. Katzan, B. Lapin, PROMIS GH (Patient-reported outcomes measurement information system global health) scale in stroke a validation study, *Stroke* **49** (2018), 147–54.

[6] F. Fischer, C. Gibbons, J. Coste, J.M. Valderas, M. Rose, A. Leplège, Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany, *Quality of Life Research* **27** (2018), 999–1014.