DEBATE

HOW SHOULD WE USE THE VISUAL ANALOGUE SCALE (VAS) IN REHABILITATION OUTCOMES? I: HOW MUCH OF WHAT? THE SEDUCTIVE VAS NUMBERS ARE NOT TRUE MEASURES

Franco Franchignoni, MD[1], Fausto Salaffi, MD[2] and Luigi Tesio, MD[3]

From the [1]Salvatore Maugeri Foundation, IRCCS, Scientific Institute of Genova Nervi, [2]Department of Rheumatology, Politechnic University of the Marche, Ancona, [3]Department of Biomedical Sciences for Health – Chair of Physical and Rehabilitation Medicine, Università degli Studi, and Department of Neurorehabilitation Sciences, Istituto Auxologico Italiano, IRCCS, Milan, Italy. E-mail: franco.franchignoni@fsm.it

We read with interest the commentary by Kersten et al. "The use of the visual analogue scale (VAS) in rehabilitation outcomes" (1), in which they explain why data from a VAS should be treated as coming from an ordinal scale. We agree with them on this point; continuity in measurement cannot be assured simply by the graphic continuum of a line, but requires continuity of the grading "from less to more" that it represents. Unfortunately, the VAS line leads one's mind to make discrete decisions on the "placement" of the measure (2). In fact, the "continuity" appears to be nothing else than the randomness surrounding each decision. The action of arriving at a rating is better conceptualized as an attempt to construct meanings influenced by, and with reference to, a range of internal and external factors and private implications, rather than as a task of matching a distance or number to a discrete internal stimulus (3). Having said that, we would like to extend the discussion, and suggest that investigators, particularly in fields such as Physical and Rehabilitation Medicine (PRM), should think twice before selecting a VAS as an outcome measure for clinical use (particularly when used as a stand-alone tool).

In our opinion, the VAS has a number of serious drawbacks, which can be classified into two main domains: (*i*) practical (such as acceptability and feasibility) and metric limitations; and, even more troublesome, (*ii*) ontological limitations.

Practical and metric limitations

The VAS appears to be a very simple metric ruler, but in fact it is not a true linear ruler from either a pragmatic or a theoretical standpoint. Among the criteria that investigators should apply to evaluate candidate outcome measures for any specific clinical trial, we have acceptability, feasibility and precision (4). Acceptability indicates how acceptable it is for respondents to complete the questionnaire (in terms of administration time, response rate, and so on), whereas feasibility refers to the ease of administration and processing (i.e. burden arising from the use, including for example the professional expertise required to apply and interpret the instrument) (5). Guyatt et al. (6) compared a VAS with a Likert scale for assessing health-related quality of life, and found that patients viewed the VAS as harder to understand and required a longer overall time (including instructions and eliciting of patient-specific information) to

complete it. Accordingly, a recent review comparing numerical rating scales (NRS), verbal rating scales (VRS) and the VAS for assessment of pain intensity in adults showed that the VAS demonstrates a lower compliance in terms of the number of patients who are able to perform the ratings, the number of correct answers, and error rate percentages (7). This happens particularly in persons of higher age and cognitive impairment, as well as in low-literacy or visually impaired patients (2).

Moreover (and in part counter-intuitively), the time to process the VAS results can be relatively long, because each physical mark placed by the patient on a VAS has to be specifically measured in terms of distance from the origin. Furthermore, this procedure introduces a risk of random errors in measuring the distance to the mark on the line.

Precision is the accuracy of the measure. This can be thought of as its capacity to discriminate across very distinct values consistently over repeated measurements (8). Streiner & Norman (2) wrote *"It is reasonable to presume that the upper practical limit of useful levels on a scale can be set at seven... the "one in a hundred" precision of the VAS is illusory; people are probably mentally dividing it into about seven segments".* Along the same lines, Jensen measured the pain intensity with a 101-NRS and found that little information was lost if the scale was coded as an 11-point scale (9). In healthy adults there is very little gain in precision with more than 7 options and hardly any above 9, in line with previous works on human information processing capacity, suggesting that 7 ± 2 levels are the finest degrees of perceptual discrimination humans can make in any situation (10). Indeed, some people with special needs are often unable appreciably to discern between more than 5 categories as indicating different levels of a variable (11). Accordingly, no significant advantage was found in the responsiveness of health-related quality-of-life measures for respiratory function when comparing a VAS with a 7-point Likert format (5).

Overall, from this point of view the VAS does not appear to show an appropriate balance between the effort needed for collecting and processing its data and the real accuracy of the measure. VAS could simply "appear" to offer more precision than other response methods, but in spite of this appearance there is no evidence that it is so (2).

Ontological limitations

A fundamental criticism that should be made with respect to the VAS approach is that it provides information on "how much", but it does not tell us exactly "how much of what".

The problem of the nature ("ontology") of the VAS has seldom been referred to in the clinical literature, which has tended to be far more interested in its practical and metric limitations, and has overlooked this fundamental measurement issue (12, 13). In our opinion, there is no getting round the ontological criticism when the VAS is used as a single-item scale (such as for pain, perceived exertion, fear of falling, etc.). For example, when faced with a single-item pain VAS, different patients might choose the same "tick" (say, 6 on a 0–10 scale), but there is no way to clearly estimate whether they mean pain intensity, peak vs average intensity, pain as a function of time, its disabling impact, the unpleasantness of the pain type, or whatever. Thus, the number of potential latent variables revealed by the rating scale is infinite and left to the examiner's interpretation.

With two (A and B) items, the "what" answer still remains vague. As for factor analysis, 3 or more items are needed to define a construct/dimension. With at least 3 items, unidimensionality can be tested, looking at the consistency of the hierarchy of item difficulty levels, according to a sample-free theoretical model. This is as old as Aristotelian logic: if item A is more difficult than item B, and item B is more difficult than item C, then item A must be more difficult than item C. Modern psychometrics, mostly through rigorous Rasch modelling, has added a probabilistic tolerance to this "fundamental measurement" axiom (13). If the syllogism is satisfied, A, B and C become ticks that mark a shared continuum "from-less-to-more difficulty".

Translation of a VAS into a NRS does not solve the problem. Whatever their number, the levels of rating scales work best if each of them reflects a clearly operationalized concept, not an anonymous numeral (14). Of course, eliciting decisions across fully operationalized alternatives requires training programmes to ensure raters' consistency. This way, a good fit to the Rasch rating scale model can often be obtained just with 3–5 well-defined ordinal categories (11, 14).

In conclusion, both the VAS and NRS suffer from some vagueness in their construct definition, the VAS further showing linearity and precision, which are fully illusory.

Genuine rating scales, with fully "operationalized" items and categories, should be preferred, and validation of the instrument should complement, not try to substitute for, a thorough conceptual definition of the variable under study and a rigorous instrument development process (15).

REFERENCES

1. Kersten P, Küçükdeveci A, Tennant A. The use of the visual analogue scale (VAS) in rehabilitation outcomes. J Rehabil Med 2012: 44: 609–610.
2. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press, 2008, p. 41–55.
3. de C Williams AC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. Pain 2000; 85: 457–463.
4. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998; 2: i–iv, 1–74.
5. Küçükdeveci A, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. J Rehabil Med 2011; 43: 661–672.
6. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987; 40: 1129–1133.
7. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011; 41: 1073–1093.
8. Tesio L. Outcome measurement in behavioural sciences: a view on how to shift attention from means to individuals and why. Int J Rehabil Res 2012; 35: 1–12.
9. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986; 27: 117–126.
10. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst) 2000; 104: 1–15.
11. Franchignoni F, Giordano A, Michail X, Christodoulou N. Practical lessons learned from use of Rasch analysis in the assessment of outcome measures. Port J Phys Rehabil Med 2010; 19: 5–12.
12. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? J Rehabil Med 2012; 44: 97–98.
13. Andrich D. Understanding resistance to the data-model relationship in Rasch's paradigm: a reflection for the next generation. J Appl Meas 2002; 3: 325–359.
14. Wolfe EW, Smith EV Jr. Instrument development tools and activities for measure validation using Rasch models: part I. J Appl Meas 2007; 8: 97–123.
15. Wolfe EW, Smith EV Jr. Instrument development tools and activities for measure validation using Rasch models: part II. J Appl Meas 2007; 8: 204–234.