Sheffield Health Economics Group

Visual Analogue Scales: do they have a role in the measurement of preferences for Health States?

Christopher McCabe and Katherine Stevens

Sheffield Health Economics Group, ScHARR, University of Sheffield

Corresponding author:

Christopher McCabe

Sheffield Health Economics Group

The University of Sheffield,

Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

Email: c.mccabe@sheffield.ac.uk

Tel 00 (44) 114 222 5454

Visual Analogue Scales: do they have a role in the measurement of preferences for Health States?

Abstract

Health state preference data are increasingly used to inform national health care resource allocation decisions. In such circumstances it is important to be confident that the data we provide to decision makers is fit for purpose. Whilst there are many unresolved issues in health state preference measurement, there are some areas of agreement on procedures that are inappropriate. The past ten years have seen the publication of a number of papers reporting substantive problems with the use of visual analogue scales to value health states. In this journal Torrance and colleagues reviewed this literature and concluded that Visual Analogue Scales have a limited but useful role in health state preference measurement. In this paper we critically review the arguments advanced by Torrance and colleagues and argue that it is increasingly clear that Visual Analogue Scales are not an appropriate method and that the time has come to accept that VAS should not be used for health state preference measurement.

Background

The results of health state preference measurement play an increasingly central role in public health care resource allocation decision making processes. [1][2][3][4] In such circumstances it is important to have confidence that the data we provide to decision makers is fit for purpose. Health state preference data is produced through a process of addressing five specific questions:

- How should we describe health?
- How should we measure health?
- How should we value health?
- Whose values should we use?
- How should values be aggregated?

In establishing whether health state preference data are fit for purpose, one has to consider the way in which each of these questions has been answered in the process of its production.[5] The current state of knowledge does not provide a consensus on the 'correct' answer to any of these questions. However, there is increasing agreement that certain answers are incorrect; for example it is no longer considered acceptable to use doctors' values of patients' health states to inform resource allocation decisions. [1][4][6] In this paper we argue that the state of knowledge regarding the Visual Analogue Scale data has reached the point that its use in health state preference measurement should now be viewed as an 'incorrect' answer to the question 'How should we value health?'.

Visual Analogue Scales

The visual analogue scale (VAS) was originally developed to allow the measurement of individual's responses to physical stimuli, such as heat.[7] VAS have been a mainstay of health state preference measurement for over 30 years and is central to the development of health state preference weights for

many of the most widely used health related quality of life measures; including the Quality of Well Being Index, the EQ-5D and the Health Utilities Indices.[8][9][10][11]

A VAS consists of a line on a page with clearly defined end points, and normally a clearly identified scale between the two end points. When VAS is used to obtain health state preference data, individuals are instructed to place the most preferred health state (e.g. full health) at one end of the line and the least preferred health state at the other end of the line. The health states for valuation are then located at points on the line that correspond to how much better and/or worse than the other health states they believe each one to be.[12]

The single greatest appeal of the VAS for health state valuation is its simplicity of use. In their review of the health state valuation literature, Brazier et al reported that completion rates for VAS based surveys were typically in excess of 90%.[7] However, there is now a substantial body of research on the limitations of VAS as a means of health state preference measurement. This literature has been reviewed in detail elsewhere, and so we will only present the key problems here.[7]

It is increasingly accepted that VAS health state preference data are subjected to a number of biases; notably context and spreading bias, and end-point aversion.[10][18] In addition, many researchers question whether there is a stable relationship between VAS health state valuation data and choice based valuation data such as Standard Gamble (SG) and Time Trade Off (TTO).[13][14] The manner in which VAS has been used by Torrance and colleagues, and other researchers, assumes such a stable relationship exists. Some researchers go as far as arguing that VAS data do not capture preferences at all.[15][16][17] Finally, some researchers question whether VAS is as simple for the respondent as is often assumed.[18]

In reviewing this literature, Torrance and colleagues argued that there was still a limited but important role for VAS in health state preference measurement. They stated that VAS "can, however, play a useful role if used with caution in

conjunction with the SG or possibly the TTO." (p333). They went on to identify the following roles for VAS: [18]

1. It can be used to familiarize the respondents with the health states and to obtain ordinal preferences;
2. It can be used to obtain cardinal health state preferences for some health states that are not subsequently measured with SG or TTO.
   a. In such circumstances, a study specific power curve should be developed to adjust the VAS scores for those states not measured by the SG or the TTO. (p333)
   b. It is unclear whether it is better to debias the VAS scores prior to the estimation of the power curve or not. (p333).

In this paper we critically review Torrance et al.'s recommendations for the use of VAS, and argue that the problems with VAS are so substantial, and its advantages so limited, that its use in health state preference measurement should be abandoned.

Using VAS to obtain ordinal health state preferences.

Ordinal health state preference data provides information on whether the respondent prefers Statse A to State B, or vice versa. They do not provide any information on the strength of the preference for State A versus B. The definition of VAS provided by Torrance et al, makes it clear that with VAS data, the distance between States A and B, has some meaning; i.e. it is intended to provide cardinal preference data. Ordinal preference data can be obtained through simple ranking exercises; i.e. where the respondent is asked to place health states in order to represent whether State A is preferred to State B; but not how much better State A is preferred to State B. The United Kingdom EQ-5D, SF-6D and UK Health Utilities Index Mark 2 valuation studies have all successfully used such ranking exercises to obtain ordinal preference data to familiarise the respondents with the health states. [23 24 25] These data have subsequently been used to estimate health state valuation

models. [19] [20] Whilst VAS data undoubtedly provide ordinal health state preferences, it is not obvious that VAS offers anything that is not provided by the ranking exercise. If the respondent is instructed that the distance between the states on the VAS should represent their strength of preference between health states,[21] [22] it seems likely that the respondent burden will be significantly increased using VAS as a prop, compared to the simple ranking exercises used elsewhere.

Using VAS to obtain cardinal health state preferences for states that are not subsequently valued using SG or TTO

The use of VAS to obtain cardinal health state preferences directly is generally deemed necessary when the number of health states to be valued is too large to allow the use of the SG or TTO technique.[22] The point at which the SG or the TTO cease to be feasible has not been established. The UK HUI2 valuation obtained 8 SG valuations from each respondent; the SF-6D study obtained 6 SG valuations from each respondent and UK EQ-5D study obtained 13 TTO valuations from each respondent.[23] [24] [25]

Whilst Torrance et al suggest that the maximum number of states to be valued simultaneously on a single VAS is in the region of 6;[18] many studies have valued considerably more. [26] [27] [28] [29]

The Canadian HUI2 valuation study required over 20 separate health state valuations from each respondent.[28] This was necessary because the study was designed to estimate individual multi-attribute utility functions (MAUF) for each respondent, in addition to a population MAUF. In studies designed to estimate population utility functions, factorial survey designs, such as that utilised in the HUI3 and UK EQ-5D study, allow the number of health states valued by each respondent to be reduced by increasing the number of respondents.[26] [24] Therefore, with the exception of research to construct individual MAUFs, respondent burden does not create a case for the use of VAS in health state preference measurement.

Against this background, we recognise that many researchers may find the perceived convenience of the VAS compared to SG and TTO a sufficient justification. To quote Torrance et al "VASs are attractive because they are simple, quick to administer and lend themselves to self completion." (p333). [18] In the next section, we consider the evidence base for Torrance et al.'s recommendation when VAS is being used to collect cardinal health state preference data.

Estimating a study specific power function

Torrance and colleagues recommend the estimation of a study-specific power function, and not the use of power functions estimated on other datasets. [1] They note that whilst it had "originally been hoped that a single consistent power function would be found between VAS scores and SG scores in all studies…it was not to be." (p332)

They argue for a power curve on two grounds; first they note that both the (adjusted) VAS and SG scales anchor at dead=0.0 and healthy=1.0. Thus any systematic relationship between VAS and SG would have to pass through 0 and 1. The power function is a simple function that meets these requirements. Secondly, they argue that the power curve has been consistently required in applied research even when the data have been adjusted for end-aversion bias or context bias. [26] [27]

With regard to the first point, there are a number of other functional forms that can be constrained to pass through the points 0 and 1; including the quadratic and cubic functions. [29] [30] [31] With regard to the second point, there is now some empirical evidence suggesting that linear, quadratic and cubic functions (all passing through the points 0 and 1) can provide a better fit than the power function for VAS-SG data, both in individual and person-mean data. [31] [29]

---

[1] Unfortunately this is quite frequently what is done in practice.[1]

Considering that the power function has not been found to be robust across datasets; it seems reasonable to consider whether the power function is robust between health states within a single dataset. This is a central assumption of the use of VAS-SG mapping functions and yet to our knowledge this issue has not been explored in the literature.

In other places it has been argued that the power function is appropriate on a theoretical basis; i.e. it reflects a constant relative risk attitude. This argument defines VAS data as values; i.e, preferences under conditions of certainty and SG data as utilities; where utilities represent preferences under conditions of uncertainty.[30]

Bleichrodt and Johannesson tested the proposal that VAS elicited points on a measurable value function and whether VAS and SG values were linked by a constant relative risk attitude.[32] Their analyses rejected both hypotheses. Schwartz subsequently found that Bleichrodt and Johannesson's data were consistent with the VAS eliciting values on a measurable value function, once the VAS data had been adjusted to take account of context bias using the Range-Frequency (RF) model.[33] However, Schwartz did not test whether the power function fitted the adjusted data.[34] Robinson et al, in a separate dataset, also found that VAS data that had been adjusted using the Range-Frequency model were consistent with VAS eliciting points on a measurable value function. [35] When they attempted to map the revised data on to SG utilities for the same states, they found that the power curve was a poor specification for the relationship.

It is worth noting, in passing, that when the RF model is applied to VAS data where the highest ranked state is given a value of 1, the adjusted value for this state will also be 1, irrespective of the value given to the parameter w.[2] The effect of this is to produce data that are no longer on the 0-1 scale required for estimating utility weights in the QALY model and which cannot be

---

[2] w is the parameter determining the relative weight given to the range and frequency effects in the range frequency model. It ranges between 0 and 1.

translated on to this scale using a linear transformation because the true adjusted value of the upper anchor of the scale is unknown.

The solution to this problem is to apply a linear transformation to the VAS data prior to the application of the RF model; an obvious candidate transformation being to map the VAS data on to a 0-100 scale. The RF model can then be applied to the transformed data. A second linear transformation can then be applied to express the adjusted (unbiased values) on the desired 0-1 scale.

If VAS data are not adjusted for RF bias, then any power function estimated on the unadjusted data will also be biased, because the effect of the RF bias is not constant across the range. To date, none of the published health state valuation models that have utilised the power curve transformation, have applied the RF model adjustment to the VAS data prior to model estimation. Therefore, it is reasonable to assume that these models have been estimated on biased data, and will in turn produce biased estimates.

Torrance and colleagues identify end-point bias as an additional consideration when using VAS to measure health state preferences.[18] It is unclear what, if any interaction there might be between RF and end-point bias, or whether a simple adjustment for end-point bias followed by the application of the Range Frequency model would adequately remove the bias from the data.

The use of VAS to obtain cardinal health state preference data commits the researcher to undertaking a number of adjustments to the data: the RF model, end-point bias adjustment, and mapping to utilities; prior to estimating a health state utility model. Simplicity in VAS data collection, we would argue, is outweighed by complexity and uncertainty in its utilisation.

Summary

Considering the experience of three substantial health state valuation studies in the UK,[23 24 25] we doubt whether VAS offers anything to the elicitation of ordinal health state preferences, over and above simple ranking exercises.

The same studies indicate that the number of health states that can be successfully valued using choice based methods, such as the SG or TTO, is such that respondent burden is unlikely to create a real need to use VAS to measure health state preferences. Further, in circumstances where the number of states to be valued is large, we would argue that individual respondent burden can be managed through the application of factorial survey designs, without sacrificing the use of choice based valuation methods, as long as the objective is to estimate population rather than individual valuation functions. Finally, the evidence is increasingly strong that VAS data are subject to bias, and the methods for adjusting for this bias have not been resolved, and even when concerns over bias have been set aside, we would argue that we do not know the functional form of the relationship between these 'Values' and the preferences that we are actually trying to measure. All we can say with confidence is that the power function is a poor model.

For these reasons, we would argue that it is time to acknowledge that VAS do not have a useful role in health state preference measurement.

References

[1] National Institute for Clinical Excellence Guide to methods for appraisal NICE London April 2004

[2] CCOHTA Guidelines for economic evaluation of pharmaceuticals: Canada. 2nd Edition Ottowa: The Canadian Coordinating Office for Health Technology Assessment; 1997

[3] PBAC Guidelines for the Pharmaceutical Industry on preparation of submission to the Pharmaceutical Benefits Advisory Committee

[4] Gold MR Siegel JE Russell LB Weinstein MC (eds) Cost effectiveness in Health Care and Medicine Oxford University Press 1996 Oxford

[5] Dolan P. The measurement of health related quality of life for use in resource allocation decisions in health care in Culyer AJ Newhouse JP Handbook of Health Economics Elsevier Science BV 2000

[6] Chapman RH, Stone PW Sandberg EA Bell C Neumann PJ A comprehensive league table of cost-utility raios and a sub-table of "Panel-worthy" studies Medical Decision Making 2000;20:451-67

[7] Brazier J Deverill M Green C Harper R Booth A. A review of the use of health status measures in economic evaluation Health Technology Assessment 1999;3(9)

[8] Kaplan RM Health Outcome models for policy analysis Health Psychology 1989;8(6):723-735

[9] Torrance GW Feeny D et al A multi attribute utility function for a comprehensive health status classification system: Health Utilties Mark 2 Medical Care 1996

[10] Feeny D Furlong W Torrance GW et al   Multi-attribute and single attribute utility functions for the Health Utilities Index Mark 3 system Medical Care 2002;40(2):113-128

[11] EuroQOl Group. EuroQol – A new facility for the measurement of health-realted quality of life. Health Policy 1990;16:199-208

[12] Torrance GW Measurement of health state utilities for economic appraisal: a review Journal of Health Economics 1986;5:1-30

[13] Bleichrodt H Johannesson M An experimental test of a theoretical foundation for rating scale valuations Med Dec Making 1997;17:208-216

[14] Dolan P Sutton M Mapping visual analogue scale health state valuations on to standard gamble and time trade off values Soc. Sci Med 1997;44(10):1519-1530

[15] Nord E Cost value analysis in health care: making sense out of QALYs Cambridge University Press 1999 Cambridge

[16] Robinson A Dolan P Williams A   Valuing health states using VAS and TTO: what lies behind the numbers Soc Sci med 1997;45:1289-97

[17] Morris J Durand A Category rating methods: numerical and verbal scales Mimeo University of York, York 1989

[18] Torrance G.W. Feeny D. Furlong W. Visual Analogue Scales: Do they have a role in the measurement of preferences for health states? Med. Dec. making 2001;21:329-334

[19] Salomon, J., A., June 2004. The use of ordinal ranks in health state valuations. IHEA Conference. USA, San Francisco.

[20] McCabe C. Brazier JE. Gilks P. Tsuchiya A. Roberts J. O'Hagan A. Stevens K. Estimating population cardinal health state valuation models from individual ordinal (rank) health state preference data Health Economics and Decision Science DP 04/01 University of Sheffield 2004

[21] Torrance, G.W., 1986. Measurement of health state utilities for economic appraisal: a review. Journal of Health Economics 5, 1- -30.

[22] Furlong W Feeny D Torrance GW Barr R Horsman J Guide to design and development of health state utility instrumentation CHEPA Working Paper #90-9 McMaster University Hamilton, Ontario 1990

[23] McCabe, C., Stevens, K., Roberts, J., Brazier, J.E., 2003. Health state values for the HUI2 descriptive system: results from a UK survey. Health Economics (forthcoming)

[24] Dolan, P., 1997. Modelling valuations for EuroQol health states. Medical Care 35 (11), 1095- -1108.

[25] Brazier, J.E., Roberts, J., Deverill, M., 2002. The estimation of a preference based measure of health from the SF-36. Journal of Health Economics 21 (2), 271- -292.

[26] Feeny D Furlong W Torrance GW Goldsmith CH Zhu Z DePauw S et al Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 System Medical Care 2002; 40(2):113-128

[27] Robinson A Loomes G Jones-Lee M Visual analogue scales, standard gambles and relative risk aversion Med Decision Making 2001;21:17-27

[28] Torrance GW Feeny DH Furlong WJ Barr RD Zhang Y Wang Q Multi-attribute utility function for a comprehensive health status classification system: Health Utilities Mark Index 2 Medical Care 1996;34(7):702-722

[29] Dolan P Sutton M Mapping visual analogue scale health state valuations on to standard gamble and time trade off values Soc. Sci Med 1997;44(10):1519-1530

[30] Dyer JS Sarin RK Relative risk aversion Mgmt Sci 1982;28(8):875-886

[31] Stevens K McCabe C Brazier JE Mapping between Visual Analogue Scale and Standard Gamble: results from a UK study using UK HUI2 valuation data HEDS Discussion Paper 04/03 University of Sheffield Sheffield

[32] Bleichrodt H Johannesson M An experimental test of a theoretical foundation for rating scale valuations Med Dec Making 1997;17:208-216

[33] Parducci A Wedell DH The category effect with rating scales: number of categories, number of stimuli and method of presentation J. of Exp. Psychology 1986;12(4):496-516

[34] Schwartz A. Rating Scales in Context Medical Decision Making 1998;18:236

[35] Robinson A Loomes G Jones-Lee M Visual analogue scales, standard gambles and relative risk aversion Med Decision Making 2001;21:17-27