The visual analogue scale for the measurement of pain is not linear

We read with interest the paper by Myles and Urquhart entitled "The linearity of the visual analogue scale in patients with severe acute pain" appearing in *Anaesthesia and Intensive Care* in the February 2005 issue[1]. Valid and effective methods for measuring pain are vital for outcomes research into pain management given the absence of objective measurement. However, we dispute Myles and Urquhart's interpretation of their data being supportive of linearity of the visual analogue scale (VAS).

They reach their conclusion principally because the average baseline pain score is almost exactly twice the average pain score when pain has subjectively halved. But does this really mean the VAS is linear? Arguably it does for the average case, but what really matters is whether the scale is linear for the majority of cases. Ironically, Myles and Urquhart have cited Bland and Altman on the point of transforming data to normalize a distribution, but they should consider Bland and Altmanís more famous work on measurement agreement[2]. The problem herein of comparing two measures on mean score alone without considering the data distributions is potentially misleading, and is often seen in repeatability or measurement comparison studies. The comparison should be made by determining the range of measurement over which 95% of the data are interchangeable (Bland and Altman limits of agreement)[2]. Contained within Figure 1 of the Myles and Urquhart paper are the raw scores for baseline and halved pain measurements. We estimated these from the figure and typed them into an Excel spreadsheet. The halved pain score was then doubled and compared to baseline score by Bland and Altman limits of agreement analysis. While the mean difference between measures was only 0.14 units, the standard deviation of the differences was 21.9 units. Therefore the 95% limits of agreement (± 1.96 times the standard deviation of the differences) were ± 43.0 units. This means that a range of 86 units is required to cover 95% of the possible agreement between baseline and halved pain scores. This is almost the entire VAS (0 to 100 units)! Clearly the VAS does not perform as a linear scale across this population.

For those unfamiliar with this type of analysis, we can demonstrate the non-linearity simply with ratios of baseline pain to halved pain. Again, looking at the ratio for each individual case and grouping them into which ratio they were closest to—1x, 1.5x, 2x, 2.5x, 3x and 3.5x—we can see that less than one-third of cases (7/22) were closest to 2x. Baseline pain to halved pain ratios were closer to 1.5x for six cases, 2.5x for four cases, 3x for three cases, 3.5x for one case, and 1x for one case. Clearly, these subjects did not use the VAS linearly.

However, this method of determining linearity by comparing two points on a scale is simplistic. An alternative approach that provides a powerful insight into scale structure is Rasch analysis. This gives the probability of selecting a particular response category in terms of the interaction between "response severity" and subject measure (in this case, pain) through an iterative logistic process. The values of response categories (in this case, VAS scores) are not assumed, so discontinuities at any stage of the scale can be detected and repaired. This makes Rasch analysis particularly useful in the development of

both single- and multi-item instruments for measuring health-related outcomes. Therefore it has been used in patient-centred outcome measurement across the breadth of medical research[3], including pain measurement[4]. Rasch analysis can also be applied to existing scales, including single item scales, to transform a non-linear scale into a linear measure. This has been done previously in pain measurement for both the Faces Scale and the visual analogue scale[5,6]. When Thomee et al applied Rasch analysis to the VAS in women with Patellofemoral Pain Syndrome they also concluded that it was not used as a linear scale. However, they were able to rescale the VAS to transform it into a true linear measure. This approach is applicable to not only the VAS, but to all single-item and multiple-item measures of pain, and should be considered whenever pain measurement is used in outcomes research.

K. PESUDOVS
Centre for Clinical Research Excellence in Ophthalmology Outcomes Research
M. J. CRAIGIE
Department of Anaesthesia and Pain Management
G. ROBERTON
Department of Anaesthesia and Pain Management,
Flinders Medical Centre and Flinders University, South Australia

References

1. Myles PS, Urquhart N. The linearity of the visual analogue scale in patients with severe acute pain. Anaesth Intensive Care 2005; 33:54-58.
2. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999; 8:135-160.
3. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) questionnaire: development and validation. Optom Vis Sci 2004; 81:769-777.
4. Boeckstyns ME. Development and construct validity of a knee pain questionnaire. Pain 1987; 31:47-52.
5. Noble BA, Loh RSK, MacLennan S et al. Comparison of Autologous Serum Eye Drops With Conventional Therapy In A Randomised Controlled Crossover Trial For Ocular Surface Disease. Br J Ophthalmol 2004; 88:647-652.
6. Thomee R, Grimby G, Wright BD, Linacre JM. Rasch analysis of Visual Analog Scale measurements before and after treatment of Patellofemoral Pain Syndrome in women. Scand J Rehabil Med 1995; 27:145-151.

The visual analogue scale for the measurement of pain is not linear—Reply

We thank Dr Pesudovs and colleagues for their comments. Accurate estimation of population parameters from a (study) sample requires representative sampling and use of appropriate statistical analyses. Measurement characteristics of a "severe pain" population, based on mean pain scores and their standard deviation, with the precision of the estimate described by the 95% confidence interval of the mean, are appropriate for the purpose of our study.

The Bland-Altmann approach, with which we are familiar, was designed for a different purpose: to determine whether two measurement methods agree or one has sufficient repeatability. Variability in this approach—the limits of agreement—is dependent on the variability of each method of measurement and the sample size. We can't imagine why a clinician would want to measure pain by halving a pain state with analgesia and then doubling the score. The Bland-Altmann approach was not designed to evaluate linearity.

The suggestion to categorize ratio pairs weakens the power of the analysis because it ignores the numerical scale of the original measurement. In any case, it is reasonable to expect a normal distribution of the resultant pairs, such that a small proportion will have extreme values: this is a characteristic of most biological variation. Rasch analysis has a related purpose: it is primarily used to convert categorical or ordinal data into a linear scale. It may well be used to analyse pain data, particularly when derived from a variety of scales, but it was not the purpose of our study.

P. S. MYLES
N. URQUHART
Alfred Hospital, Melbourne, Victoria