Correspondence to: Dr Neil R Miller, Wilmer Eye Institute, Johns Hopkins Hospital, 601 North Broadway, Maumenee B-109, Baltimore, MD 21287 9204, USA

Correspondence to: nrmiller@jhmi.edu

- 1 Felix R, Hofstetter W, Cecchini MG. Recent developments in the understanding of the pathophysiology of osteopetrosis. *Eur J Endocrinol* 1996;134:143-56.
- 2 Lehman RA, Reeves JD, Wilson WB, et al. Neurological complications of infantile osteopetrosis. *Ann Neurol* 1977;2:378-84.
- 3 Turano AF, Fagan KA, Corbo PA. Variations in clinical manifestations of osteopetrosis. *J Pediatr* 1954;44:688-94.
- 4 Chen YR, Breidahl A, Chang CN. Optic nerve decompression in fibrous dysplasia: indications, efficacy, and safety. *Plast Reconstr Surg* 1997;99:22-30.
- 5 Li KK, Lucarelli MJ, Bilyk JR, et al. Optic nerve decompression for compressive neuropathy secondary to neoplasia. *Head Neck Surg* 1997;123:425-9.

CORRESPONDENCE

The London handicap scale

Jenkinson *et al* show that London handicap scale scores are about the same if items are weighted equally rather than using the published scale weights.¹ We reached much the same conclusion using the data from which the scale weights were derived.²

Handicap is defined as disadvantage in role performance due to impairments or disabilities, which implies valuation of the extent to which role performance is affected. Value—from the viewpoint of health services research—is quantified as the “utility” of a state of health. The scale weights were derived by asking various population samples to value different combinations of problems, described using the handicap dimensions and items from the handicap scale. These were analysed to determine the contribution of each of the component parts of the description.

The fact that equal weighting gives roughly the same scores as the empirically derived weights is probably because the items were carefully chosen on the basis of clinical experience to be approximately equally spaced across the range of possible severity.

Does it matter if different weighting methods lead to much the same results? Weighting processes are inexact, be they empirically derived or equal weighting, but the second approach simply increases the level of approximation. The 95% confidence intervals around the agreement between estimated and measured scores were about ± 10 (on a 0-100 scale). This measurement imprecision arises because rating health states is difficult, leading to random measurement error, and the modelling assumed that overall valuation of a state of health could be estimated by the sum of the component parts of the description, which is almost certainly an oversimplification (although goodness of fit statistics for the model were reasonable). The London handicap scale is primarily an epidemiological tool—that is, it is intended for use in groups (such as in a clinical trial). If scores are calculated for individual patients—for example, in clinical practice—there is a further approximation, that between the values and opinions of that individual, and

“average” views of the population from which the values were derived. There is some evidence that the handicap dimensions have general validity, and there is some consensus on the values assigned to states of handicap. As part of the revision process of the *International Classification of Impairments, Disabilities and Handicaps*³ qualitative studies established strong core transcultural agreement on six domains of participation with potential to be affected by health conditions, and these corresponded to the handicap dimensions (Martin Prince, personal communication). Furthermore, a comparison of values given to states of health by Hong Kong Chinese showed good agreement with those estimated by using the London handicap scale weights (derived from Londoners).⁴ Neither was there convincing between population variation in scale weights assigned in the original scale development work.² It is not safe, however, to assume that there are no between person differences.

We are pleased to see a further independent validation of the London handicap scale. If simplification makes the scale more useful then we welcome it. The additional burden in applying the weights, however, is no more than that of adding six lines of commands in a statistical computer program (for instance, using SPSS syntax). As we have empirically derived estimates of valuations of handicap states, we see no reason why the further approximation of equal weighting is necessary.

ROWAN H HARWOOD

Health Care of the Elderly, University Hospital, Nottingham NG7 2UH, UK

SHAH EBRAHIM

Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK

- 1 Jenkinson C, Mant J, Carter J, et al. The London handicap scale: a re-evaluation of its validity using standard scoring and simple summation. *J Neurol Neurosurg Psychiatry* 2000;68:365-7.
- 2 Harwood RH. *The development, application, and evaluation of a handicap measurement scale*, [MD Thesis]. Cambridge: University of Cambridge, 1996.
- 3 World Health Organisation. *International classification of impairments, disabilities and handicaps*. Geneva: WHO, 1980. (Revised taxonomy (ICIDH-2) is at <http://www.who.int/msa/mnh/ems/icidh/index.htm> (accessed 1st April 2000).)
- 4 Lo R, Harwood R, Woo J, et al. Cross-validation of London handicap scale in Hong Kong Chinese. *Clinical Rehabilitation* 2000; (in press).

Mant *et al* reply:

Whether one decides to use the original weighting scoring system of the London handicap scale or the simpler unweighted scoring system as we proposed is essentially a trade off between the advantages of each approach. As argued by Harwood and Ebrahim, the weighting system will provide what at least seems to be a more accurate estimate of the value of a health state for only limited additional analysis. However, it should be noted that weighted systems more often than not give the appearance of greater accuracy when in fact all they produce is different results.¹ Further, even if the weights are regarded as the “gold standard” the increase in accuracy is small, and there is scope for confusion (different weights have been published),^{2,3} and the derivation of the score is more complex, which might lead to computational errors. However, perhaps the greatest advantage of the unweighted scoring system is that there is a simpler relation

between the questionnaire responses and the final score, which makes interpretation more straightforward. There is a role for both methods of analysing the London handicap scale. Use of the simpler unweighted scoring would be likely to increase uptake and use of the instrument, without any important loss of accuracy. On the other hand, researchers who are already familiar with the weighting system have little to gain from switching to the unweighted system.

JONATHAN MANT

Department of Primary Care and General Practice, University of Birmingham, B15 2TT, UK

CRISPIN JENKINSON

Health Services Research Unit, University of Oxford, UK

JUDY CARTER

DERICK WADE

Rivermead Rehabilitation Centre, Oxford, UK

SIMON WINNER

Department of Clinical Geratology, Radcliffe Infirmary, Oxford, UK

- 1 Jenkinson C. Why are we weighting? *Soc Sci Med* 1991;32:1413-16.
- 2 Harwood R, Rogers A, Dickinson E, et al. Measuring handicap: the London handicap scale, a new outcome measure for chronic disease. *Quality in Health Care* 1994;3:11-16.
- 3 Harwood RH, Ebrahim S. *Manual of the London handicap scale*. Nottingham: University of Nottingham, 1995.

Mass volume measurement in severe head injury

We read with interest the article by Stocchetti *et al* on the accuracy and feasibility of the ellipsoid and the Cavalieri method in assessment of the volume of intracranial mass lesions in patients with severe head injury.¹ We agree with the authors that the volume of intracranial lesions, and its change over time, is important in the diagnosis and management of patients with head injury and in the evaluation of clinical trials.²

However, the methodology used in the study raised our concern. We have several comments on their statements, because they are potentially misleading.

(1) The statement that computer based reading of mass lesions is the choice when accurate volume estimation is necessary, is insufficiently founded. Tracing CT lesions on a digitised screen automatically calculating area and hence volume, is a hazardous task: delineating hyperdense and hypodense lesions from normal surrounding intracranial structures cannot always be performed reliably, due to isodensity of normal brain tissue at some edges and due to partial volume effects. Moreover, lesion tracing is the same as area estimation using a simple device such as a point counting grid with sufficient grid points, and is in fact not superior at all.³

(2) Volume estimations based on Cavalieri's principle have to fulfill one absolute requirement: randomness.^{4,5} The volume of any object may be estimated from randomised and parallel sections separated by a known distance by summing up the areas of all cross sections of the object and multiplying this sum by the known intersection distance. The total area of all cross sections may be estimated by a stereological point counting method.³ A systematic array of grid intersection points is superimposed on each section. Giving random positioning of the test array on each section, the total number of grid intersection points hitting the object of interest affords an unbiased estimate of the

total area. In the study of Stocchetti *et al* randomness has most probably not been accounted for, as it has not been mentioned in the text and as the grid has not randomly been placed onto the CT slices.¹

(3) When applying Cavalieri's principle it becomes mathematically possible to calculate the coefficient of error of the individual lesion mass volume. It declines in direct proportion to the total number of CT planes and to the total number of grid intersection points. Generally, a coefficient of error of less than 0.05 is obtained if the number of CT planes used is 10 or more, and the number of grid intersection points is 50 or more.^{3,5} From our own practical experience we and others know that CT at standard settings (5 or 8 mm slice thickness) almost never renders sufficient sections through the lesion mass, except for very huge subdural or extradural haematomas. Spiral CT with a 3 mm section distance may overcome this problem. Another advantage of Cavalieri's principle is its applicability to any mass lesion irrespective of size and form.

(4) The average difference between the applied technique and the reference computer based value is 0.57 (SD 9.99!) ml for the Cavalieri method and 0.20 (SD 15.48!) ml for the ellipsoid method, suggesting on average acceptable agreement. However, what really matters is the accuracy, validity and reliability of the individual volume measurements. That these are not very high can be derived from the huge standard deviations of the average differences and from the considerable limits of agreement in the graphical depiction of the results.

Accuracy of the individual measurements has to be high as in the trauma coma data bank (TCDB) classification a volume of greater than 25 ml is defined as a mass lesion.²

(5) Although three examiners read the scans, the interobserver variability was calculated with analysis of variance (ANOVA) on the mean volumes. No intraobserver variability studies were conducted which can be considered as an omission.

(6) The TCDB CT classification being the resultant of the status of the mesencephalic cisterns, the degree of midline shift, and the presence of a mass lesion provides a ranking order of the severity of the initial injury (I=normal, II=diffuse injury, III=diffuse injury with swelling, IV=diffuse injury with shift, V=operated mass lesion, and VI=non-operated mass lesion). Intracranial mass lesion volume, although important, is but one of the measured indices in the TCDB classification. We showed recently that the TCDB CT classification system for patients with severe head injury has in fact a high interobserver and intraobserver reliability when used by clinicians without special training in neuroradiology.⁶

PIETER E VOS

OSCAR J M VOGELS

Department of Neurology, University Medical Centre
St Radboud, PO Box 9101, 6500 HB Nijmegen, the
Netherlands

TJEMME BEEMS

Department of Neurosurgery

HENK O M THIJSSSEN

Department of Neuroradiology

CEES ZIMMERMAN

Department of Intensive Care

Correspondence to: Dr Pieter E Vos
p.vos@czoneu.azn.nl

- 1 Stocchetti N, Croci M, Spagnoli D, *et al*. Mass volume measurement in severe head injury: accuracy and feasibility of two pragmatic methods. *J Neurol Neurosurg Psychiatry* 2000;68:14-7.
- 2 Marshall LF, Eisenberg HM, Jane JA, *et al*. A new classification of head injury based on computerised tomography. *J Neurosurg* 1991;75: S14-20.
- 3 Gundersen HJG, Jensen EB. The efficiency of systematic sampling in stereology and its prediction. *J Microsc* 1987;147:229-63.
- 4 Cavalieri B. *Geometria degli indivisibili (original edition, Bologna, 1635)*. Torino: Unione Tipografico-Editrice, 1966:1-320.
- 5 Vogels OJM, Zijlmans JCM, van 't Hof MA, *et al*. MR volume estimation of subcortical brain lesions and ventricular cerebrospinal fluid: a simple and accurate stereologic method. *AJNR Am J Neuroradiol* 1995;16:1441-5.
- 6 Vos PE, van Voskuilen AC, Beems T, *et al*. Evaluation of the Trauma Coma Data Bank CT Marshall classification for traumatic brain injury. *J Neurotrauma* (in press).

Stocchetti and Colombo reply:

We are grateful to Vos *et al* for their comments. Because the main reason for our research was the fact that intracranial mass lesions are rarely measured, it is reassuring that some centres have documented expertise on such measurement.

We do not think, however, that our data, and the conclusions drawn from them, are potentially misleading, and we will try to clarify our arguments.

As indicated in the paper, we compared two pragmatic methods and a computer based method. There are, of course, limitations to each method, and tracing on the computer screen can be tricky; however, a careful tracing is feasible in expert hands and we think that the resulting calculation still gives a very acceptable reference point. If this reference method is to be questioned, an appropriate, preferably superior, method should be identified. We cannot think of any pragmatic method that would be the best choice.

Regarding the other points that aroused the concern of Vos *et al*, we agree on many and will try to clarify them.

(1) Randomness is an important prerequisite; it was not mentioned in the paper but the grid was placed on the CT slices at each reader's convenience and choice. Whether this was random enough can be debated, but it seems to ensure an adequate guarantee against systematic error.

(2) Our data did not obscure the beauties of the Cavalieri method. The direct estimator performed better than the Ellipsoid method, particularly for irregularly shaped lesions. We agree that the method performs even better with bigger lesions reconstructed using thin slices. As one of our goals was to describe feasibility, however, the comments of Vos *et al* further stress that the best can be obtained from the Cavalieri method at the price of more time and work, as we verified and reported in the paper. Counting more than 50 points in more than 10 slices adds to the precision, but seriously increases the burden of measurements.

(3) We did not base our comparison only on mean data. We share the concerns of Vos *et al* on the analysis of mean data to the extent that we have used another method, based on that of Bland and Altman. This method compares every single lesion, obtained with each method, against the corresponding reference value. The results of this detailed comparison are illustrated in figs 1 and 2 of our paper. Numerical data, summarising the analysis according to Bland and Altman, are reported in the text and table. Both in the results sec-

tion and in the discussion we stated that the mean data were not able to describe the discrepancies found in single cases.

(4) Accordingly, we assessed interobserver variability by ANOVA of individual measurements, and not of the mean data. In other words, we asked whether the measurement of any specific lesion by one examiner was significantly different from the other examiners' results.¹ The ANOVA on the readings by the three examiners using the Ellipsoid method gave a p value of 0.86, and the same analysis applied to the Cavalieri direct estimator gave a p value of 0.81; we therefore concluded that this analysis excluded significant differences. It seems that the intraobserver variability was not omitted, but was in fact logically considered, and we must apologise if the text was not clear.

(5) We agree with the final paragraph of Vos *et al* on the structure of the Marshall classification, in which volume is one part of the grading. That was correctly indicated in our paper. From our experience in multicentre, international clinical trials, we are less optimistic about the proper application of the TCDB CT classification, but that is another point in favour of improving the methods for CT readings.

In conclusion we have applied a methodology that seems solid enough to substantiate our conclusion and, we hope, to fulfil the requirements of careful and competent readers.

NINO STOCCHETTI
ANGELO COLOMBO

Neuroscience Intensive Care, Policlinico Hospital
IRCCS Milano, Via Francesco Sforza 35, 20122
Milano, Italy

Correspondence to: Dr N Stocchetti

1 Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994;13:2465-76.

BOOK REVIEWS

Neurosurgery. The Scientific Basis of Clinical Practice. Third edition. Edited by CROCKARD, HAYWARD, and HOFF (two volumes, £285.00). Published by Blackwell Science, Oxford, 1999. ISBN 0632048387.

This two volume book is unique in providing a comprehensive overview of all the aspects of basic neuroscience relevant to the neurosurgeon. As such it can be regarded as an essential source of information for neurosurgical trainees around the world. The authors state in the foreword to this edition that their ambition was to "reflect in as up to date fashion as possible the expanding knowledge so essential both to trainees and established neurosurgeons alike if they are to base their clinical practice on a proper scientific basis." They can be congratulated on fulfilling their objective with such a well presented and beautifully illustrated new edition. Although clearly intended for a neurosurgical readership, there are sections that would be valuable for practitioners in other disciplines, particularly neurologists, oncologists, and orthopaedic surgeons involved in spinal surgery.

The third edition has expanded with the increased range of knowledge required by the