global issues and open avenues for high-quality submissions.

Ana Marusic
Matko Marusic
Editors in Chief
Croatian Medical Journal
Zagreb, Croatia

References

1. Western medical journals and the 10/90 problem [editorial]. *CMAJ* 2004;170(1):5.
2. Marusic A, Marusic M. Small scientific journals from small countries: breaking from a vicious circle of inadequacy [editorial]. *Croat Med J* 1999; 40(4):508-14.
3. Marusic M, Marusic A. Good editorial practice: editors as educators [editorial]. *Croat Med J* 2001;42(2):113-20.
4. Marusic A, Misak A, Kljakovic-Gaspic M, Marusic M. *Educatione ad excelentiam* — ten years of the *Croatian Medical Journal*. *Croat Med J* 2002;43(1):1-7.
5. Marusic A, Marusic M. What can medical journal editors do in war? *Lancet* 2002;360(Suppl):s59-60.
6. Marusic A, Marusic M. Teaching students how to read and write science: a mandatory course on scientific research and communication in medicine. *Acad Med* 2003;78(12):1235-9.

DOI:10.1053/cmaj.1040034

Grading evidence

Holger Schünemann and his colleagues in the GRADE Working Group[1] have taken an important first step in developing a universally acceptable grading system for denoting the quality of evidence and associated recommendations. Current systems use alphabetic, numeric or colour-coded nominals or ordinals, which represent discontinuous, qualitative (or at most semiquantitative) and hence imprecise categories.[2,3] In many instances the resulting imprecision exceeds that of the information it attempts to convey. I would like to suggest an alternative.

The most informative and sophisticated form of measurement uses a continuous scale with consistent intervals.[2] Interval scales are common in clinical medicine, being used for measurements of blood pressure, temperature, heart rate and weight, and for nearly all laboratory measurements. The consistent intervals allow values to be combined as averages and deviations. If the scale starts at zero, it becomes a ratio scale, which allows ratio statements such as "twice as big" or "half as much."[2]

My suggestion is to implement a system already in common use throughout the world: the 100-interval ratio scale, which is widely used for currencies and for grading performance and which is based on the most common counting practice, the decimal system.

The use of this scale to grade the quality of scientific measurements would not be new. Statistical confidence limits around a point estimate are expressed as percentages, as are the sensitivity, specificity and predictive values of diagnostic interventions.[2,4] Probabilities and likelihoods may be expressed on the 100-interval scale or can be readily converted to it, while utility, the relative value of alternative choices, is also often expressed as a value out of 100.[4] Perhaps most importantly, Bayes' factors, the mathematical expression of how disease indicants modify diagnostic hypotheses, fit well with the 100-interval scale and may also be used for therapeutic interventions.[4-6] This scale even facilitates the use of odds, as odds to the base 100 are equivalent to percentages.[6]

The disadvantage of the system is that it may give a sense of precision that does not exist. For example, clinicians

primarily use subjective degrees of belief in diagnostic reasoning. Indeed, even objective observations are expressed with confidence intervals, not just as point estimates. However, a 100-interval ratio scale seems preferable to a system of only 4 grades that do not fit either the Bayesian type of reasoning used in clinical practice or the clinical decision analysis that is increasingly recommended for use in complex clinical and policy problems.[7]

Gerald J.M. Tevaarwerk
Professor Emeritus
University of Western Ontario
Zutphen, the Netherlands

References

1. Schünemann HJ, Best D, Vist G, Oxman AD, for the GRADE Working Group. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations [editorial]. *CMAJ* 2003;169(7):677-80.
2. Hassard TH. *Understanding biostatistics*. St. Louis: Mosby Year Book; 1991.
3. Upshur REG. Are all evidence-based practices alike? Problems in the ranking of evidence [editorial]. *CMAJ* 2003;169(7):672-3.
4. Dawson-Saunders B, Trapp RG. *Basic and clinical biostatistics*. Norwalk (CT): Appleton & Lange; 1994.
5. Tevaarwerk GJM. Measuring the efficacy and cost-effectiveness of laboratory tests. *Ann R Coll Phys Surg Can* 1995;28(6):217-20.
6. Tevaarwerk GJM. A Bayesian approach to treatment efficacy: therapeutic likelihood ratios and odds change values. *Ann R Coll Phys Surg Can* 1998;31(7):319-26.
7. Lusted LB. *Introduction to medical decision-making*. Springfield (IL): CC Thomas; 1968.

DOI:10.1053/cmaj.1031694

[The authors respond:]

We agree with Gerald Tevaarwerk that 100-interval scales are useful for many purposes. However, for several reasons (some of which Tevaarwerk has listed and some we discussed[1]), we disagree that such scales would be useful in the setting of guidelines and recommendations for grading evidence.

First, interval scales suggest a degree of precision that does not exist in the evaluation of quality of evidence; the types of study designs that can be used to determine quality of evidence are limited, which results in few categories. Second, an interval scale for quality of evidence would suggest that we can express quality of evidence in terms of multiples (e.g., "twice as much quality"), but we do not believe that this interpretation is justified. Third, with regard to presentation and practicality, an interval scale would present challenges in the production and dissemination of guidelines. Fourth, with regard to Bayesian reasoning in clinical practice or clinical decision analysis, guideline panels provide guidance to clinicans when they make their recommendations and assign letters, numbers or symbols to those recommendations. Ideally, guideline developers consider clinical decision analysis before they make their recommendations, and we see no use for interval scales that describe the quality of evidence in decision analysis.

Holger J. Schünemann
Department of Clinical Epidemiology and Biostatistics
McMaster University
Hamilton, Ont.
Dana Best
Division of General Pediatrics
Children's National Medical Center
Washington, DC
Gunn Vist
Andrew D. Oxman
Department of Health Services Research
Norwegian Directorate for Health and Social Welfare
Oslo, Norway