

Are all evidence-based practices alike? Problems in the ranking of evidence

Ross E.G. Upshur

§ See related article page 677

Evidence-based medicine is an integral component of undergraduate medical curricula, postgraduate training and clinical practice. The concepts of levels of evidence and grades of recommendations are central to the definition of evidence-based practice, for they attempt to standardize and provide clinicians with cogent rules to appraise published research, determine its validity and summarize its utility in clinical practice.

The strategy of separating levels of evidence and grades of recommendations originated in preventive care and was a defining feature of what became the Canadian Task Force on Preventive Health Care.¹ It reinforces the notion that evidence exists in a hierarchical fashion, with some study designs being more subject to bias than others and therefore possibly providing weaker justification for clinical decisions.

Appreciation of differences in the quality of evidence has been followed by a proliferation of evidence hierarchies. *CMAJ* recently published osteoporosis guidelines that used one such variation.² The existence of multiple classifications for evaluating and structuring evidence and the differing interpretations of grades of recommendations on the

basis of this evidence pose potential problems for clinicians. I have identified at least 4 evidence hierarchies.¹⁻⁴ Considerable differences exist between them, both in respect to what counts as the highest quality evidence and what constitutes the strongest possible recommendation (Table 1). Further differences occur as one proceeds down the hierarchy.

What is at stake in this analysis of differing evidence hierarchies? First, the concept of evidence-based medicine was founded in part to reduce unnecessary inconsistencies and encourage standardized practice. The belief that a standardized approach to evidence decreases variation is one of the tenets of evidence-based practice. However, the various hierarchies are differentially permissive or restrictive with respect to what counts as best evidence. As a consequence, the inconsistent nomenclature introduces a wide range of possibility for “evidence-based” practices.

Second, the hierarchies give differential weight to consensus and evidence, with some allowing and others not allowing consensus to be included in their assessments. Processes of evidence assessment that rely on consensus in making recommendations introduce an opaque dimension to how the recommendations are made and compromise

Table 1: Highest level of evidence and strongest grade of recommendation in 4 evidence hierarchies

Source of evidence hierarchy	Highest level of evidence for a treatment or intervention	Conditions for a grade A recommendation
Canadian Task Force on Preventive Health Care ¹	1 = at least one RCT	A study (including meta-analysis or systematic review) that meets all design-specific criteria well
Scientific Advisory Council of the Osteoporosis Society of Canada ²	1+ = systematic review or meta-analysis of RCTs 1 = one randomized trial with adequate power	Supportive level 1 or 1+ evidence plus consensus
Centre for Evidence-Based Medicine ³	1a = systematic review with homogeneity of RCTs 1b = individual RCT with narrow confidence interval 1c = all or none*	Consistent level 1 studies
Scottish Intercollegiate Guidelines Network ⁴	1++ = high-quality meta-analyses, systematic reviews of RCTs, or RCTs with very low risk of bias 1+ = well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with low risk of bias 1- = meta-analyses, systematic reviews of RCTs, or RCTs with high risk of bias	At least one meta-analysis, systematic review or RCT rated as 1++ and directly applicable to the target population; or a systematic review of RCTs or a body of evidence consisting principally of studies rated as 1+ directly applicable to the target population and demonstrating overall consistency

Note: RCT = randomized controlled trial.

*The Centre for Evidence-Based Medicine explains “all or none” as follows: “met when *all* patients died before the treatment became available, but some now survive on it; or when some patients died before the treatment became available, but *none* now die on it.”

the objectivity that clinicians demand of their evidence. If extra-evidential considerations are part of the deliberation, it is not clear to the critical mind how this occurred.

Why does it matter? Consider the following example. A pharmaceutical company distributed evidence-based guidelines to my clinic. According to the guidelines, level 1 (best) evidence required one well-designed randomized control trial, and a grade A recommendation required one level 1 study. The guidelines were accompanied by a report of the randomized trial sponsored by that same company and published in a peer-reviewed journal. By disseminating the guidelines with the supportive paper, the company sought to persuade me that I would be following evidence-based guidelines if I prescribed the drug.

Recently, there has been acknowledgement of a distinction between evidence-based practitioners and evidence users.⁵ Particularly in primary care, there is a trend toward using pre-appraised sources of evidence. One of the most popular of these sources is InfoPOEMs (patient-oriented evidence that matters), a daily email service that provides summaries of research studies relevant to primary care (www.infopoems.com). Each summary is accompanied by the level of evidence, using the Oxford Centre nomenclature. Such knowledge transfer is welcome, but as similar services proliferate it is unclear which nomenclature will be used and how it will affect practice. Is your evidence 1++, 1+, 1a or level 1? Do we understand each other when we say 2c or 2+D?

Clearly, the proliferation of evidence hierarchies and grades of recommendations is intended to promote the use of evidence-based approaches in health care and provide clinicians with a guide to reliable knowledge. However, it also threatens to introduce confusion and devalue the currency of evidence-based medicine.

What are the possible solutions? First, I think journal editors should collectively insist on standardization with respect to the use of evidence hierarchies, particularly in the dissemination of clinical practice guidelines. I think agreement on one nomenclature and one system of recommendations within journals would be welcome. This would be similar to requiring standardized abstracts or standardized reporting requirements for randomized trials (e.g., CONSORT). Agreement on what constitutes an evidence hierarchy and a grade of recommendation would lead to international harmonization of understanding what an evidence-based recommendation is and allow more consistent patterns of practice, to the benefit of the patient.

Second, the process by which this agreement is achieved should include strong input from the very creators of evidence-based medicine. Because the movement had its birth in Canada, it is important for institutions such as McMaster University to take the lead in harmonizing these hierarchies. Because evidence-based medicine is now an international movement, perhaps a consensus conference devoted to creating an international standard is required.

Third, I think clinicians in the field should also be involved in the decision process, to facilitate evidence transfer and uptake. Their input in terms of how these hierarchies should be constructed and what would indicate meaningful credibility in evidence is needed.

Finally, research is required to determine the utility and acceptability of different formats for presenting evidence ratings and grades of recommendations. Ironically, the creation of these classifications has not as yet been informed by research but is driven in large part by expert opinion. This is what makes the article by Schünemann and colleagues from the GRADE Working Group timely (see page 677).⁶ The group is conducting a systematic evaluation of the strengths and weaknesses of different means of communicating levels of evidence and grades of recommendations. It has the laudable goal of “reaching agreement on a common, sensible approach to grading quality of evidence and strength of recommendations.” However, the group’s findings indicate a paucity of research to guide its work. The challenges ahead for proponents of evidence hierarchies are significant, but continued proliferation of evidence hierarchies and grades of recommendations potentially dilutes much of what evidence-based medicine seeks to achieve. It is hoped that the GRADE Working Group will be inclusive in its process of seeking consensus on these matters and find ways of making even more explicit and transparent how such ratings are assigned.

This article has been peer reviewed.

Dr. Upshur is Director of the Primary Care Research Unit and Assistant Professor in the Departments of Family and Community Medicine and of Public Health Sciences and in the Joint Centre for Bioethics, University of Toronto, Toronto, Ont.

Competing interests: None declared.

Acknowledgements: Dr. Upshur is supported by a New Investigator Award from the Canadian Institutes of Health Research and a Research Scholar Award from the Department of Family and Community Medicine, University of Toronto.

References

1. Canadian Task Force on Preventive Health Care. History and methods. Available: www.ctfphc.org (accessed 2003 Aug 27).
2. Brown JP, Josse RG; Scientific Advisory Council of the Osteoporosis Society of Canada. 2002 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada. *CMAJ* 2002;167(Suppl 10):S1-34.
3. Centre for Evidence-Based Medicine. Levels of evidence and grades of recommendation. Oxford: The Centre. Available: www.cebm.net/levels_of_evidence.asp (accessed 2003 Aug 27).
4. Wright PJ, English PJ, Hungin AP, Marsden SN. Managing acute renal colic across the primary-secondary care interface: a pathway of care based on evidence and consensus. *BMJ* 2002;325:1408-12.
5. Guyatt GH, Meade MO, Jaeschke RZ, Cook DJ, Haynes RB. Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills [editorial]. *BMJ* 2000;320:954-5.
6. 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. *CMAJ* 2003;169(7):677-80.

Correspondence to: Dr. Ross E.G. Upshur, Rm. E349B, 2075 Bayview Ave., Toronto ON M4N 3M5; fax 416 480-4536; rupshur@idirect.com