

The Evidence-based Medicine Paradigm: Where are We 20 Years Later? Part 2

Shashi S. Seshia, G. Bryan Young

ABSTRACT: In Part 2, we discuss the challenges of keeping up with the ‘literature,’ evidence-based medicine (EBM) in emerging economies and the Neurosciences, and two recent approaches to classifying evidence. We conclude by summarizing information from Parts 1 and 2 which suggest the need to critically re-appraise core elements of the EBM paradigm: (1) the hierarchical ranking of evidence, (2) randomized controlled trials or systematic reviews as the gold standard for all clinical questions or situations, (3) the statistical tests that have become integral to the ‘measurements’ for analyzing evidence, and (4) re-incorporating a role for evidence from basic sciences and pathophysiology. An understanding of how cognitive processes influence clinical decisions is also necessary to improve evidence-based practice. Emerging economies may have to modify the design and conduct of clinical research to their settings. Like all paradigms, EBM must keep improving with input from the grassroots to remain beneficial.

RÉSUMÉ: Paradigme de la médecine fondée sur des preuves : où en sommes-nous 20 ans plus tard? Deuxième partie. Dans cette deuxième partie, nous discutons des défis que présentent le suivi de la littérature, la médecine fondée sur des données probantes (MFDP) dans les économies émergentes et les neurosciences ainsi que deux approches novatrices de classification des données. Nous concluons par un sommaire de l'information contenue dans les parties 1 et 2 qui suggère que nous devons réévaluer de façon critique des éléments fondamentaux du paradigme de la MFDP : 1) le classement hiérarchique des données ; 2) les études randomisées contrôlées ou les revues systématiques comme étalon or en ce qui concerne toutes les questions ou situations cliniques ; 3) les tests statistiques qui sont devenus une partie fondamentale des « mesures » utilisées pour analyser les données et 4) le rôle des données des sciences de base et de la physiopathologie qui devrait être rétabli. La compréhension de la façon dont les processus cognitifs influencent les décisions cliniques est également nécessaire pour améliorer la pratique médicale fondée sur des preuves. Les économies émergentes pourraient devoir modifier la conception et la réalisation de la recherche clinique dans leur contexte. Comme tout paradigme, la MFDP doit s'améliorer constamment en tenant compte des contributions du milieu pour demeurer bénéfique.

Can J Neurol Sci. 2013; 40: 475-481

How do we keep up with the literature?

The era of formalized evidence-based medicine (EBM) coincided with the widespread use of computers, development of software, and computerised medical informatics. There has also been an exponential increase in the literature. In their paper titled, “Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?”,¹ the authors draw attention to the flood of information that deluges us daily. Their solution and that of others,² is “systematic reviews,” (SRs) and “pre-appraised” information bases and summaries; in addition they exhort us to “burn your traditional textbooks,” for “a new breed...evidence-based...services,” so that “clinicians will enjoy a set of increasingly accessible sources of evidence...that acknowledge the most current EBM thinking...”.^{2,3} In 1997, Feinstein and Horwitz lauded the suggestions to have sources of authority such as the Cochrane Collaboration develop authoritative appraisals; at the same time, they presciently listed the biases that could compromise quality of these appraisals.⁴ Their cautionary note is justified by the potential and actual errors and limitations in the literature (including Cochrane reviews), discussed in Part 1.

Despite the CONSORT (Consolidated Standards of Reporting Trials) group’s hope that “investigators who conduct inadequate trials...should not be able to pass through the publication process without revelation of their trials’

inadequacies,”⁵⁻⁷ they have. Deficiencies occur in the text, and in abstracts from which most readers obtain their information.⁸⁻¹²

The reporting of studies assessing diagnostic tests, are subject to shortcomings similar to those for randomized controlled trials (RCTs).¹³ The STARD (Standards for Reporting of Diagnostic Accuracy) guidelines were meant to rectify such deficiencies,¹⁴ but improvements have been slow.¹⁵ The GRADE (Grading of Recommendations Assessment, Development, and Evaluation) working group has developed recommendations for grading evidence for diagnostic tests.¹⁶ However, these provide limited advice on biostatistical aspects.¹⁷ Also, guidance is almost exclusively for “dichotomous diagnostic tests,” failing to address the grey zones for radiological and electrophysiological tests, so commonly used in practice.

From the Department of Pediatrics, Division of Pediatric Neurology (SSS), University of Saskatchewan, Royal University Hospital, Saskatoon, Saskatchewan; Department of Clinical Neurological Sciences (GBY), Western University, London, Ontario, Canada.

RECEIVED OCTOBER 5, 2012. FINAL REVISIONS SUBMITTED JANUARY 3, 2013.
Correspondence to: Shashi S. Seshia, Department of Pediatrics, Division of Pediatric Neurology, University of Saskatchewan, Royal University Hospital, 108 Hospital Drive, Saskatoon, Saskatchewan, S7N 0W8, Canada. Email: sseshia@yahoo.ca.

Montori and Guyatt have drawn attention to “corrupted reports,” and advise “users” of medical literature to strengthen the art of critical appraisal and exercise healthy skepticism,³ however vaunted the source of information.

Recent approaches to assessing evidence

Two groups, GRADE and the Oxford Centre for Evidence Based Medicine (OCEBM) have attempted to improve upon earlier classifications of levels and recommendations for evidence. There are similarities in, and some differences between their approaches.

The GRADE working group was formed around 2004 to address shortcomings in the then systems for grading evidence and recommendations,¹⁸ which were discussed in Part 1. Some of the major principles of GRADE are out-lined in Table 1 (Part 2). Detailed discussion is beyond the scope of this review. Clinicians are directed to a series of papers;^{16,18-24} another series for SRs, health technology assessments, and clinical practice guidelines is published in the *Journal of Clinical Epidemiology*.²⁵ GRADE has been adopted widely,^{21,26} and is being used in the Neurosciences.^{27,28} At least three authors have expressed reservations about aspects of GRADE; their critiques include, (i) the hierarchy (ranking) of evidence is maintained; (ii) as with the earlier EBM paradigm, GRADE is still opinion, not scientific based, and (iii) GRADE is complex to use with inter-observer agreement even between those trained or experienced in SRs being variable across several GRADE domains.²⁹⁻³¹

The 2011 OCEBM levels and methods are comprehensive (Part 2, Table 2),³² and takes a stepwise (Steps 1 through 5) and hierarchical (Levels 1 through 5) approach that is specific for the clinical question. Systematic reviews are rated Step 1 (Level 1) for most questions; GRADE also advocates SRs as the basis for judging evidence, but discourages the incorporation of SRs in levels of evidence (see Table 1 Part 2). The OCEBM uses similar criteria as GRADE to upgrade or downgrade evidence. The information in Table 2 (Part 2) also provides a succinct list of the various methods for seeking evidence for several clinical questions. To our knowledge, information from the basic sciences and pathophysiology are not incorporated in either method. The 2011 OCEBM levels of evidence should not be

used without the introductory document, both freely downloadable from the website.³²

Neither GRADE nor the 2011 OCEBM levels of evidence have been subject to critical review for their ‘effectiveness,’ but both incorporate a philosophy of continuous improvement.

EBM, treatment trials, and developing economies

The core principles of ethical authoritative clinical practice are universal. However, there are unique challenges for developing economies. These include:

1. EBM texts strongly suggest the use of concepts of p values confidence intervals (CIs), Number needed to treat (NNT) etc from RCTs when discussing treatment of specific disorders with our patients.² This is difficult to do when patients and families are not well educated.
2. The ethical challenges of applying EBM practices and conducting trials in Sub-Saharan Africa have been discussed by Oricha and Yauri.³³ Their experience will reflect that in many countries. The uncertainty principle can be challenging or impossible to apply, and randomization difficult to implement; the authors suggested that other methods such as sequential trials, may be better alternatives.³³
3. The risk of exploitation in the conduct of trials is greater for vulnerable populations, such as those in developing economies.³⁴

Therefore, developing economies need to develop EBM models specific for their respective populations.

EBM and the Neurosciences

The cornerstones of EBM are as firmly entrenched in the Clinical Neurosciences as in all other disciplines, directly and indirectly, impacting treatment, investigations and management of both neurological conditions and frequently associated comorbidity.³⁵⁻³⁸

Brigo suggested that the use of placebo or a clearly inferior anti-epileptic drug (AED) in RCTs testing new AEDs was neither scientifically nor ethically acceptable, head to head comparison with the best possible treatment being the ideal.³⁹

Table 1: Key points in the GRADE system

1. Judgments about evidence and recommendations are complex.
 2. “Expert opinion” should not be classified as a category of evidence, because judgment (“expert opinion”) is necessary for interpretation of all evidence.
 3. The GRADE system does not eliminate the “inevitable need for judgments” regarding quality of evidence and strength of recommendation, and acknowledges the subjectivity in making these decisions.
 4. GRADE classifies quality of evidence into one of four levels: high, moderate, low and very low, and recommendations as either strong or weak.
 5. Judgments about the quality of evidence should be based on a systematic review of the relevant research.
 6. Systematic reviews should not be included in a hierarchy of evidence i.e., as a level or category of evidence since even a well done review may include studies of varying quality.
 7. Evidence based on RCTs without important limitations is considered high quality, that from observational studies with important limitations or without special strengths is considered low quality, while expert reports of clinical experience and case reports are classified as very low quality. However, an RCT or any other study can be downgraded because of study limitations, inconsistency of results, indirectness of evidence, imprecision, or publication bias. Conversely, large magnitude of effect, plausible confounding that would reduce a demonstrated effect and dose-response gradient would increase the quality of a study.
 8. Strength of recommendation is influenced by the quality of evidence, balance between desirable and undesirable effects, individual values and preferences, and cost.
 9. GRADE defines quality of evidence differently for systematic reviews and guidelines.
 10. For clinical guidelines: “judgments about the quality of evidence...should be made in the context of systematic reviews, such as Cochrane reviews.” “However clinical guidelines are only as good as the evidence and judgments they are based on.”
-

Please see references,¹⁸⁻²⁵ and text for details.

Table 2: OCEBM 2011 levels of evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or <i>n</i> -of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
What are the COMMON harms? (Treatment Harms)	Systematic review of randomized trials, systematic review of nested case-control studies, <i>n</i> -of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or <i>n</i> -of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Non-randomized controlled cohort/follow-up study**	Case-series, case-control, or historical controlled studies**	Mechanism-based reasoning

OCEBM: Oxford Centre for Evidence-Based Medicine. Working group= Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, Moschetti I, Phillips B, Thornton H, Goddard O, Hodgkinson M. *Level may be downgraded or upgraded, depending upon several factors.

** Systematic review is generally better than an individual study. Mechanism-based reasoning (Step and level 5): "Involves an inference from mechanisms to claims that an intervention produces a patient-relevant outcome. Such reasoning will involve an inferential chain linking the intervention (such as antiarrhythmic drugs) with a clinical outcome (such as mortality). (Howick)" Please see information on website for details.³² Reproduced with the kind permission of OCEBM.

The principles for testing AEDs suggested by Mattson et al. in 1983 are still valid,⁴⁰ and the double blind comparative RCT on AEDs published in 1985 remains a reference standard in adults.⁴¹ The comparative study between clobazam, carbamazepine and phenytoin in children remains an example for well designed pragmatic effectiveness comparative studies on AEDs in children.⁴² The RCT comparing surgery with medical treatment for temporal lobe epilepsy may well be a reference standard for studies comparing surgical and medical treatments.⁴³ The recent paper by Barnett highlights the principles for the conduct of meaningful yet ethical ("intercultural and intercontinental") trials, a model not just for clinical research in the Neurosciences, but for health care in general;⁴⁴ His key points (which complement the discussion in Part 1) are worthy of reproduction almost verbatim:

1. There must be a "burning" unsolved therapeutic question.
2. "Genuine and guarded randomization" is necessary to provide definitive answers; anecdotes are of no value and historical controls may mislead.
3. Informed consent must be taken ethically. Patients (whether in the active treatment or control group) must never be knowingly exposed to risks.
4. The natural history of the illness under question and preceding pilot observations should be incorporated in sample size calculations.
5. The treated and control groups must have identical baseline characteristics.
6. The description of the patients to be randomized must be clearly stated.

7. If there are any subgroups to be analyzed, they must be clearly described in the protocol.
8. All outcome events must be reported to the Central Office of the Trial without delay. These outcomes must be described in prose and by check-offs in the follow-up papers, assuring the reader of consistency. Failure to comply with these "cannot be tolerated".
9. The outcome details must be sent promptly to the Central Office. They will be needed to cope with the "Stopping Rules." The clinical coordinator of the Central Office must scrutinize each document from every collaborating centre, looking for gaps in obligatory details, or failure to have all reports corrected for English in each centre. Any ambiguities must be reported and discussed immediately with the participating physician in the centre; lack of conformity with the protocol must be recognized and immediately discussed with the investigators in the participating centre. Delays diminish the likelihood of accuracy of recall.
10. The protocol must contain a listing of the patients who are regarded as "Exclusions".
11. The methods planned ahead for the data analysis must be meticulously described in the Protocol. These analyses will form the gold bricks of the trial conclusions.
12. It is mandatory that the protocol, including the Methods of operation of the Trial and the methods of analyses be set out and published before the Results paper.
13. The Principal investigators, once a results paper appears, are ethically obliged to respond quickly to any queries about the conduct and the analyses of the Trial. The final data-bank

must be regarded as a transparent document available to all, once the results have been peer-reviewed and published.

14. Patients can never be dropped once randomized, and with care need never be lost.
15. Participants should be rejected who receive consulting fees from industry connected to the subject of the trial.
16. Biostatistical experts have to be involved.

The hierarchical levels of evidence may be problematic for assessing information on chronic conditions like multiple sclerosis.^{45,46} The complexity of assessing the evidence in neurocritical care, especially in the context of head injury, is illustrated by the inability of a SR to establish “the isolated benefit” of intracranial pressure monitoring in severe traumatic brain injury;⁴⁷ given the considerable between study heterogeneity, a narrative or qualitative review may have provided clinical insights from the authors’ labors (see Part 2, Appendix #1). The complexity is further highlighted by the debate surrounding the DECRA trial on decompression craniectomy in diffuse traumatic brain injury (Part 2, Appendix #2).⁴⁸⁻⁵⁷ Lemcke provides two contrasting case examples (Part 2, Appendix #3), that exemplify the challenges of generalizing results from RCTs to the individual patient.⁵³

CONCLUSIONS

In this two part review, attention has been drawn to several methodological biases, deficiencies and conflicts of interest issues in papers dealing with treatment. To quote Hirst and Altman: “Much published research across all medical specialties is not useful, may be misleading, wasteful and even harmful;” journals, publishers, and producers of SRs have a responsibility to provide clinicians and the public with reliable information.⁶ Investigators and oversight committees have an ethical responsibility to ensure that the design, conduct and analyses of trials are performed in a manner that contributes to beneficial, useful and reliable information.

Readers must neither place blind faith in any source of information nor adopt “slavish cookbook approaches;”⁴ even non-industry funded research may have biases.⁵⁸ Critical appraisal has always been a major “pragmatic and central” feature of EBM;⁵⁹ studies have shown the importance of continuous reinforcement and practice.⁶⁰

The EBM paradigm “embodies just one model or world view.”⁶¹ Several authors have questioned the hierarchical concept of evidence, opinions supported by “important shortcomings” in commonly used levels of evidence.¹⁹ Their views are best summed by complementary quotes from (i) some of Sir Michael Rawlins’ concluding remarks in his Harveian oration: “Hierarchies of evidence should be replaced by accepting-indeed embracing-a diversity of approaches,”^{62,63} and (ii) Jaddad and Enkin: “There is no best evidence except in reference to particular types of problems, in particular contexts,”³⁴ remarks relevant to the Neurosciences.

Credible authors have also questioned the unqualified proclamation of the RCT and SR as gold standards;^{4,34,62} The RCT is a powerful and often essential tool when conducted in a scientifically sound and morally ethical manner; however, the RCT is only one piece to help solve a clinical “puzzle.”³⁴ Frequently, it has to be complemented by other types of studies to get more reliable information about benefit and harm in the

individual and population at large.^{34,39,62-64} Rigorous post-marketing surveillance (phase IV trials) should be a mandatory requirement for the release of any new treatment.³⁴

The 2011 OCEBM levels of evidence (Part 2, Table 2),³² addresses some of the foregoing critiques. In addition, the stepwise approach (Part 2, Table 2) also reflects the philosophy articulated by Sackett et al: “if no randomised trial has been carried out for our patient’s predicament, we must follow the trail to the next best external evidence...”⁶⁵ Another feature of the 2011 OCEBM system is the explicit avoidance of “recommendations,” an integral component of most hierarchical methods to assess evidence (Part 1, Table 1). Rather, as espoused in the EBM paradigm, clinicians have to use clinical judgment to ensure that the external evidence applies to the individual patient.^{2,65}

The OCEBM working group might wish to consider some additional revisions for improvement: (i) drop any reference to a perceived hierarchy as reflected by the Levels 1 to 4 in their current table; this would not compromise their message, (ii) incorporate information from the basic sciences and pathophysiological principles,⁴ and (iii) explicitly state that “authoritarian” clinical expert opinion, either individual or committee, without reference to the type of evidence on which it is based, is of doubtful value and should be discouraged, a principle espoused by the American Academy of Neurology.³⁷

The complexity of statistically presented medical information is increasing. Statistics (‘measurement’), another cornerstone of the EBM paradigm, can be and are misused.⁶⁶⁻⁶⁸ Greenhalgh cautioned, “It is possible to be seriously misled by taking the statistical competence (and/or the intellectual honesty) of authors for granted.”⁶⁷ The ready availability of soft-ware, developed to meet the needs of EBM, to perform complex statistical tests, and SRs including Cochrane reviews, increases the probability of misuse; the need for guidance has been emphasized.⁶⁹ Mills’ and Greenhalgh’s remarks^{66,67} apply even more today than they did almost two decades ago. In addition, several limitations and flaws have been identified in the statistical and related methods that have become integral to the conduct of and ‘measurements’ in clinical trials; therefore, a critical re-appraisal of the methodological and statistical methods in health care, treatment trials and tests included, is urgently needed.^{68,70}

The implementation of recent recommendations to improve the credibility of industry-sponsored research,⁷¹ together with laws such as The Trials and Experimental Studies Transparency (TEST) Act in the United States, and access to unpublished company data, should help rectify existing shortcomings in the evaluation of treatments.^{72,73} The inadequate disclosure of conflict of interest by authors needs to be addressed.⁷⁴

Cognitive biases can contribute to diagnostic errors, inappropriate use of and interpretation of tests, and also influence how we interpret and apply what we read.⁷⁵⁻⁷⁹ The study and understanding of diagnostic reasoning and biases are essential to evidence-based practice.

Evidence-based medicine has ancient origins, but some of its more recent explicit concepts are new.⁶⁵ The paradigm has contributed to a more widely prevalent behavior of critical appraisal and knowledge-based practice. Several RCTs and SRs have had positive impacts on health care. Simultaneously and inevitably, limitations and deficiencies have also been identified.

The EBM paradigm must be held to the same high standards of critical review and evidence, that EBM rightly set on the practice of medicine in 1992.⁵⁹ An improved paradigm that addresses deficiencies, misuses and abuses will be an even more positive legacy.

Sir Austin Bradford Hill considered integrity and code of honour to be "second nature to those qualified in medicine,"⁸⁰ attributes not the prerogative of physicians alone. These virtues remain central to the conduct of clinical research and practice.

ADDENDUM

A recent editorial (Heneghan C, Godlee F. "Where next for evidence based healthcare?" *BMJ*.2013;346:f766) supports the arguments made in our two part review.

ACKNOWLEDGEMENTS

1) The continuing support of the Department of Pediatrics and Faculty of Medicine, University of Saskatchewan, Saskatoon, Canada, is gratefully acknowledged (SSS); (2) We thank Drs. Bill Bingham, Richard Huntsman, Haresh Kirpalani, Terry Klassen, Shoo Lee, and Molly Seshia for reviewing an earlier draft (or drafts: MS & RH), and suggesting improvements. Dr. John Baier optimized Figure 2 (Part 1) for publication; (3) SSS dedicates this review to his teachers in India and UK, and to Professor Victor Chernick, and the late Professors Bill Bowman and Harry Medovy of Winnipeg for being role models of evidence-based teaching and practice; (4) GBY dedicates this review to Drs. Henry Barnett, Warren Blume and Charles Bolton for similar reasons; (5) We are grateful to several authors (of references), too numerous to mention, who responded graciously and promptly to emails for clarification (Professor H. Heinzl, Drs. Gordon Guyatt and Rafael Perera did so on several occasions). The authors thank Professor H. Heinzl (Core unit for medical statistics and informatics, Medical University of Vienna, Vienna, Austria) for clarifying type III error and for his guidance in developing the example used to illustrate types I, II and III errors (Appendix #1, Part 1). Acknowledgements are not meant to imply endorsement by our personal reviewers or others. Any errors of omission or commission and opinions are ours alone.

Note: We have tried to limit duplication and enhance continuity between Parts 1 and 2; neither is complete without the other. The EQUATOR network in Reference 6, Part 2 is a recent addition to the several organizations that have the objective of providing resources and education to improve and monitor health research reporting. Further information can be obtained on their website: <http://www.equator-network.org/about-equator/what-we-do/>

REFERENCES

- Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010;7:e1000326.
- Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-based Medicine; 4th ed. Toronto: Churchill Livingstone Elsevier, 2011.
- Montori VM, Guyatt GH. Progress in evidence-based medicine. *JAMA*. 2008;300:1814-6.
- Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". *Am J Med*. 1997;103:529-35.

- Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152:726-32.
- Hirst A, Altman DG. Are peer reviewers encouraged to use reporting guidelines? A survey of 116 health research journals. *PLoS One*. 2012;7:e35621.
- CONSORT group. [cited 2012 Dec 13]. Available from: <http://www.consort-statement.org/>.
- Berwanger O, Ribeiro RA, Finkelsztejn A, et al. The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals. *J Clin Epidemiol*. 2009;62:387-92.
- Shukralla AA, Tudur-Smith C, Powell GA, Williamson PR, Marson AG. Reporting of adverse events in randomised controlled trials of antiepileptic drugs using the CONSORT criteria for reporting harms. *Epilepsy Res*. 2011;97:20-9.
- Assadi R, Zarghi N, Sapahri Shamloo A, Nikooiyan Y. Evidence-based abstracts: what research summaries should contain to support evidence-based medicine. *Int J Evid Based Healthc*. 2012;10:154-8.
- DeMauro SB, Giaccone A, Kirpalani H, Schmidt B. Quality of reporting of neonatal and infant trials in high-impact journals. *Pediatrics*. 2011;128:e639-44.
- Kiehna EN, Starke RM, Pouratian N, Dumont AS. Standards for reporting randomized controlled trials in neurosurgery. *J Neurosurg*. 2011;114:280-5.
- Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174:469-76.
- STARD. [Update April 2008; cited 2012 Dec 13]. Available from: <http://www.stard-statement.org/>.
- Bossuyt PM. STARD statement: still room for improvement in the reporting of diagnostic accuracy studies. *Radiology*. 2008;248:713-4.
- Schunemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106-10.
- Hess AS, Shardell M, Johnson JK, et al. Methods and recommendations for evaluating and reporting a new diagnostic test. *Eur J Clin Microbiol Infect Dis*. 2012;31:2111-6.
- Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490.
- Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res*. 2004;4:38.
- Atkins D, Briss PA, Eccles M, et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res*. 2005;5:25.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924-6.
- Guyatt GH, Oxman AD, Kunz R, et al. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336:995-8.
- Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. *BMJ*. 2008;336:1049-51.
- Guyatt GH, Oxman AD, Kunz R, et al. Incorporating considerations of resources use into grading recommendations. *BMJ*. 2008;336:1170-3.
- Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383-94.
- Thornton J, Alderson P, Tan T, et al. Introducing GRADE across the NICE clinical guideline program. *J Clin Epidemiol*. 2013 Feb;66(2):124-31.
- Nakagawa TA, Ashwal S, Mathur M, Mysore M, Committee For Determination Of Brain Death In Infants Children. Guidelines for the determination of brain death in infants and children: an update of the 1987 task force recommendations-executive summary. *Ann Neurol*. 2012;71:573-85.
- Pringsheim T, Davenport W, Mackie G, et al. Canadian Headache Society guideline for migraine prophylaxis. *Can J Neurol Sci*. 2012;39:S1-59.

29. Goldenberg MJ, Borgerson K, Bluhm R. The nature of evidence in evidence-based medicine: guest editors' introduction. *Perspect Biol Med.* 2009;52:164-7.
30. Upshur R. Making the grade: assuring trustworthiness in evidence. *Perspect Biol Med.* 2009;52:264-75.
31. Hartling L, Fernandes RM, Seida J, Vandermeer B, Dryden DM. From the trenches: a cross-sectional study applying the GRADE tool in systematic reviews of healthcare interventions. *PLoS One.* 2012;7:e34697.
32. Centre for Evidence-Based Medicine, University of Oxford. Oxford, UK. [cited 2012 Dec 12]. Available from: <http://www.cebm.net/>.
33. Oricha BS, Yauri MB. Uncertainty principle versus clinical equipoise in clinical trials in Sub-Saharan Africa: Are they really tenable? *Ann African Med.* 2003;2(2):99-100.
34. Jadad AR, Enkin MW. *Randomized controlled trials.* Oxford, UK: Blackwell Publishing BMJI Books, 2007.
35. Burneo JG, Demaerschalk BM, Jenkins ME, editors. *Neurology: An evidence-based approach.* New York: Springer 2012.
36. Wiebe S, Demaerschalk B. Evidence based care in the neurosciences. *Can J Neurol Sci.* 2002;29:115-9.
37. Gronseth G, French J. Practice parameters and technology assessments: what they are, what they are not, and why you should care. *Neurology.* 2008;71:1639-43.
38. French J, Gronseth G. Lost in a jungle of evidence: we need a compass. *Neurology.* 2008;71:1634-8.
39. Brigo F. New anti-epileptic drugs: overcoming the limits of randomised controlled trials. *Int J Evid Based Healthc.* 2011;9:440-3.
40. Mattson RH, Cramer JA, Delgado Escueta AV, Smith DB, Collins JF. A design for the prospective evaluation of the efficacy and toxicity of antiepileptic drugs in adults. *Neurology.* 1983;33:14-25.
41. Mattson RH, Cramer JA, Collins JF, et al. Comparison of carbamazepine, phenobarbital, phenytoin, and primidone in partial and secondarily generalized tonic-clonic seizures. *N Engl J Med.* 1985;313:145-51.
42. Anonymous. Clobazam has equivalent efficacy to carbamazepine and phenytoin as monotherapy for childhood epilepsy. Canadian Study Group for Childhood Epilepsy. *Epilepsia.* 1998;39:952-9.
43. Wiebe S, Blume WT, Girvin JP, Eliasziw M, Effectiveness and Efficiency of Surgery for Temporal Lobe Epilepsy Study Group. A randomized, controlled trial of surgery for temporal-lobe epilepsy. *N Engl J Med.* 2001;345:311-8.
44. Barnett HJM. Experiences with the execution of intercultural, intercontinental trials - Part I. *Can J Neurol Sci.* 2013;40:324-9.
45. Goodin DS, Reder AT. Evidence-based medicine: promise and pitfalls. *Mult Scler.* 2012;18:947-8.
46. Gronseth GS, Ashman E. The AAN response to evidence-based medicine: promise and pitfalls. *Mult Scler.* 2012;18:949-50.
47. Mendelson AA, Gillis C, Henderson WR, Ronco JJ, Dhingra V, Griesdale DE. Intracranial pressure monitors in traumatic brain injury: a systematic review. *Can J Neurol Sci.* 2012;39:571-6.
48. Cooper DJ, Rosenfeld JV, Murray L, et al. Decompressive craniectomy in diffuse traumatic brain injury. *N Engl J Med.* 2011;364:1493-502.
49. Honeybul S, Ho KM. The influence of clinical evidence on surgical practice. *J Eval Clin Pract.* 2012; May 8 [Epub ahead of print]. PMID:22568805
50. Honeybul S, Ho KM, Lind CR. What Can Be Learned from the DECRA Study. *World Neurosurg.* 2013;79:159-61.
51. Iaccarino C, Schiavi P, Servadei F. Decompressive Craniectomies: Time to Discuss Not the DECRA Study but the Comments to the DECRA Study. *World Neurosurg.* 2013;79:78-9.
52. Hutchinson PJ, Timofeev I, Koliass AG, et al. Decompressive craniectomy for traumatic brain injury: the jury is still out. *Br J Neurosurg.* 2011;25:441-2.
53. Lemcke J. Taming a tyrannosaur: the decompressive craniectomy for traumatic brain injury has to become an evidence-based procedure. *Eur J Neurol.* 2011;18:543-4.
54. Chi JH. Craniectomy for traumatic brain injury: results from the DECRA trial. *Neurosurgery.* 2011;68:N19-20.
55. Torres R. DECRA...Where do we go from here? *Surg Neurol Int.* 2012;3:54.
56. Kitagawa RS, Bullock MR. Lessons from the DECRA Study. *World Neurosurg.* 2013;79:82-4.
57. Cooper DJ, Rosenfeld JV, Wolfe R. DECRA Investigators' Response to "The Future of Decompressive Craniectomy for Diffuse Traumatic Brain Injury" by Honeybul et al. *J Neurotrauma.* 2012;29:2595-6.
58. Drazen JM. Believe the data. *N Engl J Med.* 2012;367:1152-3.
59. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA.* 1992;268:2420-5.
60. Kahlon G, Mansi IA, Banks DE. Educating medical students in evidence-based medicine: what we should expect as a starting point for our house officers. *South Med J.* 2012;105:184-8.
61. Booth A. Evidence-based practice: triumph of style over substance? *Health Info Libr J.* 2011;28:237-41.
62. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet.* 2008;372:2152-61.
63. Rawlins MD. De Testimonio: On the evidence for decisions about the use of therapeutic interventions. London, UK: Royal College of Physicians, 2008.
64. Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci.* 2011;13:217-24.
65. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312:71-2.
66. Mills JL. Data torturing. *N Engl J Med.* 1993;329:1196-9.
67. Greenhalgh T. How to read a paper. Statistics for the non-statistician. II: "Significant" relations and their pitfalls. *BMJ.* 1997;315:422-5.
68. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med.* 2009;28:3189-209.
69. Hartling L, Hamm M, Milne A, et al. Validity and inter-rater reliability testing of quality assessment instruments. Rockville (MD): Agency for health care research and policy, US department of health and human services, 2012.
70. Matcham J, Julious S, Pyke S, et al. Proposed best practice for statisticians in the reporting and publication of pharmaceutical industry-sponsored clinical trials. *Pharm Stat.* 2011;10:70-3.
71. Mansi BA, Clark J, David FS, et al. Ten recommendations for closing the credibility gap in reporting industry-sponsored clinical research: a joint journal and pharmaceutical industry perspective. *Mayo Clin Proc.* 2012;87:424-9.
72. Drazen JM. Transparency for Clinical Trials - The TEST Act. *N Engl J Med.* 2012;367(9):863-4.
73. Strech D, Littmann J. Lack of proportionality. Seven specifications of public interest that override post-approval commercial interests on limited access to clinical data. *Trials.* 2012;13:100.
74. Kesselheim AS, Wang B, Studdert DM, Avorn J. Conflict of interest reporting by authors involved in promotion of off-label drug use: an analysis of journal disclosures. *PLoS Med.* 2012;9:e1001280.
75. Crosskerry P, Nimmo GR. Better clinical decision making and reducing diagnostic error. *J R Coll Physicians Edinb.* 2011;41:155-62.
76. Crosskerry P, Abbass AA, Wu AW. How doctors feel: affective issues in patients' safety. *Lancet.* 2008;372:1205-6.
77. Crosskerry P. Context is everything or how could I have been that stupid? *Healthc Q.* 2009;12 Spec No Patient:e171-6.
78. Crosskerry P. A universal model of diagnostic reasoning. *Acad Med.* 2009;84:1022-8.
79. Crosskerry P. Perspectives on diagnostic failure and patient safety. *Healthc Q.* 2012;15 suppl:50-6.
80. Hill AB. Medical ethics and controlled trials. *Br Med J.* 1963;1:1043-9.

Appendix to Part 2

Please consider our views on the journal articles and topics (below) as constituting a journal club critical appraisal discussion; yours may differ. The numbering of references corresponds to their first citation in the text. The three examples that follow are inter-linked. Readers will need to access the references that form the basis for the discussions that follow.

1. Discussion of "Intracranial pressure monitors in traumatic brain injury: a systematic review."⁴⁷

The authors conducted a systematic review "in accordance with the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines." Six studies were ultimately included in the review.

Not unexpectedly, there was considerable between study heterogeneity examining the role of intracranial pressure (ICP) monitoring and outcome in traumatic brain injury (TBI). There were differences within and between studies in definitions of severe TBI, and selection of patients for monitoring, realities not unlike those faced in many clinical settings. Clinicians often did not adhere to Brain Trauma Foundation guidelines; this may have been due to personal bias or to individual patient variables. The authors drew attention to the complexities in the neurocritical care setting. They were unable to draw conclusions about the isolated benefit of ICP monitoring in severe TBI.

Patients who have increased ICP and head injury often have other confounding clinical variables that have to be treated in their own right. The management of raised ICP does not occur in isolation in such situations. We suggest that the complexity of critical care patients, including the heterogeneity between patients, would preclude a quantitative systematic review or make the results unreliable or generalizable. A narrative review (or clinical audit) can provide valuable practical information in such settings.

2. The DECRA trial.^{48,57}

This randomized single blinded (outcome) controlled trial was done to determine if decompression craniectomy (DC) in severe diffuse TBI with intracranial hypertension improved outcome, comparing treated group with standard-care group. The investigators found a greater risk of a worse score on the extended Glasgow coma outcome scale at six months and greater risk of an unfavorable outcome in the treated group.

Our objectives are not to reproduce the debate in the cited literature.⁴⁹⁻⁵⁶ Rather, we address five methodological issues (please refer to Part 1 for the background to these issues) that are important to assess in trials, and affect generalizability of the results. We do not have the expertise to discuss the complex statistical analyses:

(i) Enrollment bias. Of the 3478 subjects assessed for trial eligibility, only 155 were enrolled (4.5%); The reasons for rejecting the others are not clear; of the 155, 88% were from Australia or New Zealand, despite the multi-centre nature of the study (could this have added another bias?) Those who enrolled subjects and conducted the study were not blinded; only those who analysed outcome were blinded. Biases in enrollment will affect generalizability.

(ii) Sample size calculation. It is not clear what hypothesis was used in study design (see discussion in Part 1), and sample size calculations. An erroneous statistical method for the hypothesis and for any interim changes to methodology (see below) can affect reliability.

(iii) Was the trial stopped too early? The authors conducted an interim analysis, and made some changes in methodology to complete

the trial in a "reasonable time frame." They recalculated the sample size, and met it. Although they achieved the recalculated the sample size, comments from experts on the limitations of stopping trials early (discussed in Part 1) would be of value.

(iv) Clinical methodological issues. Critiques have drawn attention to selected subset of patients with TBI, unusual choice of DC procedure, and the intracranial pressure level at which intervention was undertaken, as reasons why the results cannot be generalized. The investigators of DECRA agreed that their results could only apply to patients similar to theirs.

(v) Differences in baseline characteristics (most important). The investigators acknowledged important imbalances in baseline clinical characteristics between treated and untreated group, with the treated group having more severe signs: A violation of the fundamental assumption of any study that the two groups be evenly matched for baseline characteristics. The harmful effect of DC was not evident, once this imbalance was factored in, although there was no evidence for benefit. Given the difference between the baseline characteristics of the groups (which could only have become apparent after study completion), is it reasonable to suggest that the study design and conduct would not have permitted an assessment of benefit?

3. The clinical dilemma-the individual patient.⁵³

Lemcke discussed the DECRA trial, and used two personal cases to illustrate the challenges of applying evidence from even a well designed randomized controlled trial to an individual patient, in the context of the use of DC for TBI. The cause of injury was being thrown from the fifth floor in a 16-year-old girl and a lorry accident in an 18-year-old girl respectively. Their baseline characteristics and treatments were similar. The 16-year-old died two hours after the operation. The second is now working as a hotel manager, and other than two operation scars has no sequelae. How would 'you' act in 'your' next similar case?

The examples in this appendix reflect the challenges of conducting trials and performing quantitative systematic reviews in critical care and acute neurosurgical settings such as head injury, where subtle but clinically important differences in baseline characteristics exist between patients: no two are alike.