REVIEW ARTICLE

The Evidence-based Medicine Paradigm: Where are We 20 Years Later? Part 1

Shashi S. Seshia, G. Bryan Young

ABSTRACT: The evidence-based medicine (EBM) paradigm, introduced in 1992, has had a major and positive impact on all aspects of health care. However, widespread use has also uncovered some limitations; these are discussed from the perspectives of two clinicians in this, the first of a two part narrative review. For example, there are credible reservations about the validity of hierarchical levels of evidence, a core element of the EBM paradigm. In addition, potential and actual methodological and statistical deficiencies have been identified, not only in many published randomized controlled trials but also in systematic reviews, both rated highly for evidence in EBM classifications. Ethical violations compromise reliability of some data. Clinicians need to be conscious of potential limitations in some of the cornerstones of the EBM paradigm, and to deficiencies in the literature.

RÉSUMÉ: Paradigme de la médecine fondée sur des preuves : où en sommes-nous 20 ans plus tard? Première partie. Le paradigme de la médecine fondée sur des données probantes (MFDP) introduit en 1992 a eu un impact positif majeur sur tous les aspects des soins de santé. Cependant, son utilisation répandue a également mis au jour certaines limites. Nous discutons de ces limites du point de vue de deux cliniciens dans la première partie de cet examen narratif. Il existe, par exemple, des réserves crédibles concernant la validité des niveaux hiérarchiques de preuves, un élément clé du paradigme de la MFDP. De plus, des lacunes potentielles et réelles dans la méthodologie et l'analyse statistique ont été identifiées, non seulement dans plusieurs essais cliniques randomisés qui ont été publiés, mais également dans les revues systématiques, deux sources de données très prisées pour établir les classifications dans la MFDP. Les manquements à l'éthique compromettent la fiabilité de certaines données. Les cliniciens doivent être conscients des limites potentielles présentes dans certains principes de base du paradigme de la MFDP et des lacunes présentes dans la littérature.

Can J Neurol Sci. 2013; 40: 465-474

A "new approach to teaching the practice of Medicine" described as a "paradigm shift," and called "evidence-based medicine" (EBM), was first discussed in 1992.[1] The paradigm was meant to replace an "authoritarian (opinion-based)" attitude in health-care with an "authoritative (evidence-based)" one.[1,2]

This two part narrative review addresses some issues at the core of EBM. These include hierarchical levels, the randomized controlled trial (RCT), systematic (including Cochrane) reviews (SRs) and statistical methods to assess evidence. The objective is not only to inform but also to facilitate continuing dialogue on a paradigm that is now embedded in the teaching, practice, research, writings and management of health-care,[2-5] Neurosciences included.[6-8]

Methods

The principles out-lined in Straus et al,[2] were used to seek references pertinent to EBM, first through PubMed. The initial search term "Evidence-based Medicine" yielded 79,600 results. Limits were then placed: human; guidelines; reviews; systematic reviews; meta-analyses, and editorial confined to the English language. Additional searches were made using Medical Subject Headings (MeSH), "randomized controlled trials," "confidence intervals," "number needed to treat," "CONSORT," "levels of evidence," "surrogate markers," "composite end points", and "GRADE." The resulting lists were visually scanned, and titles even remotely relevant to the search objectives were selected. In addition, searches were made using the names of those considered to have been major contributors to the EBM paradigm (Cochrane, Guyatt, Glasziou, Greenhalgh, Hanaghan, Sackett, and Straus). Further searches were made through Google Scholar; also, the websites of the Cochrane Collaboration,[9] Oxford Centre for EBM (OCEBM)[10] and CONSORT (Consolidated Standards of Reporting Trials) group[11] were visited. Often, one reference provided a link to others of interest. A total of 600 + abstracts were short-listed and the original papers obtained and read. Those included in this review were considered the most pertinent.

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. Sechia, 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.

Evolving concepts of the EBM paradigm

In 1995, Davidoff et al suggested that identifying the best evidence using epidemiological and biostatistical ways of thinking was one of five "linked ideas" of EBM.[12] In 1996, Sackett et al wrote, "evidence-based medicine is the conscientious, explicit, and judicious use of *current* best evidence in making decisions about the care of individual patients...means integrating individual clinical expertise with the best available external evidence from systematic research;" they added that "by best available external clinical evidence we mean clinically *relevant* research...especially from patient centered clinical research..." (*italics* ours).[13]

Straus and McAlister defined EBM as, "*the process* of systematically finding, appraising and using contemporaneous research findings as the basis for clinical decisions."[14] In 2009, Djulbegovic et al suggested "We should consider EBM as a continuously evolving heuristic structure for optimizing clinical practice."[4]

Levels of evidence

The hierarchy is a cornerstone of EBM. The evolution of the hierarchical approach has been discussed recently.[15] Most of the levels of evidence in the medical literature (including the American Academy of Neurology's) have been based on study design, and modelled on the 1979 Canadian task force on the periodic health examination.[15-18] Atkins et al identified at least 50 variations.[19] Typically, the levels of evidence range from I (highest) to III or IV or V (lowest), recommendations (example, for a treatment or test) being rated accordingly (Table). Hierarchies assign all non-RCT based evidence into lower levels. Evidence from the use of basic sciences such as microbiology, pharmacology, physiology, and pathology etc., that have contributed to seminal therapeutic discoveries, are not referred to in many levels of evidence.[20]

Several authors have questioned the evidence underlying the hierarchical classification.[21-31] In 2008, Sir Michael Rawlins wrote, "The notion that evidence can be reliably placed in hierarchies is illusory."[23] In addition, shortcomings were found in "prominent" levels of evidence and grading of recommendation systems being used in 2000.[19] The GRADE (Grading of Recommendations Assessment, Development and Evaluation)[32,33] system and the 2011 OCEBM levels of evidence,[10] may address some limitations of the earlier classifications of evidence and will be discussed in Part 2.

The Randomized controlled trial (RCT)

Sir Austin Bradford Hill is often credited for the modern RCT,[30] but he presciently not only warned about the "potentially dangerously misleading" nature of "poorly constructed trials," but also cautioned, "any belief that the controlled trial is the only way would mean not that the pendulum has swung too far but had come right off the hook."[34] Schwartz and Lellouch discussed the RCT, and advocated a complementary role for the then commonplace trial (which they termed "pragmatic").[35] Cochrane hoped that the RCT would "open a new world of evaluation," of medical treatment,[36] and it has. However, despite Sir Austin's cautionary note,[34] the RCT was proclaimed the gold standard in EBM.[13]

Not unexpectedly, flaws began to be identified in RCTs, raising questions about its unqualified position at the "summit" of evidence.[20-23,25,28,30,31,37-39] The reservations include: (i) Most RCTs are based on carefully selected groups of patients in a controlled setting, and provide information about efficacy, and rarely on effectiveness, i.e., actual effects on the spectrum of patients encountered in clinical practice, (ii) RCTs usually do not provide information about long-term benefit or adverse effects, (iii) they give information about the statistical "average" for the patient group; to quote Buchanan and Kean, "a patient is neither a mean nor median,"[40] and (iv) the roles for complementary well designed methods such as pragmatic trials and observational studies including those using cohort data bases, etc. have been minimized.[23,30,39]

In addition, there are many situations, where RCTs are not appropriate, and others where the clinical state of the subject, the nature of the underlying disorder and the number of confounding variables, may limit the kind of trials that can be done;[20,30,41] also,

Table: Simplified composite frame-work of commonly used Levels of Evidence

Level	Basis of evidence	Recommendation
I	One "well designed" RCT→SR of such RCTs	A
	? n-of-1 RCT	
II	Less well designed RCTs; well designed case cohort	
	studies; other well designed controlled trials	B
III	Observational and other case series, case reports	C
IV	Expert opinion; consensus	D

Most clinicians will be familiar with variations of this simplified out-line for therapeutic questions. Over time, several classifications sub-classified studies (example I-1, 2, 3; II-1, 2, 3 etc) according to their interpretation of the strength of evidence from the study type; systematic reviews (SRs) replaced randomized controlled trials (RCTs) as level 1 evidence in some. Grading of recommendation (for benefit, harm or no effect) correspondingly ranges from definitive through probable, possible and unproven (A through D).

https://doi.org/10.1017/S0317167100014542 Published online by Cambridge University Press

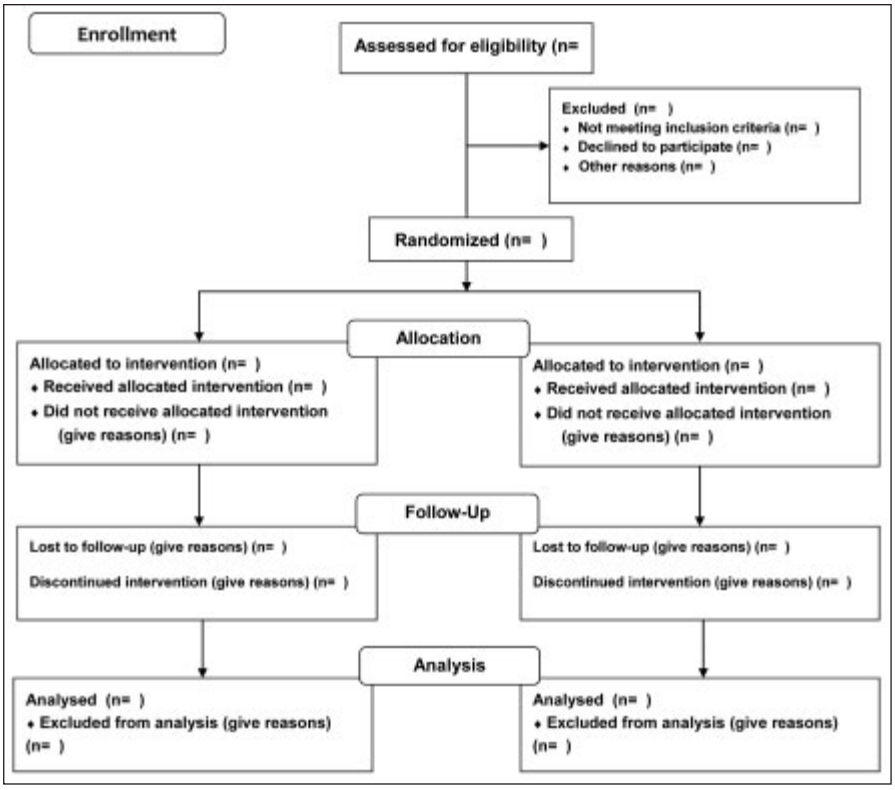

Figure 1: *CONSORT 2010 Flow Diagram. The CONSORT Statement and the CONSORT Explanation and Elaboration Document are distributed under the terms of the Creative Commons Attribution License, which permits use, distribution, and reproduction in any medium, provided the original author and source are credited. Original figure reproduced from references.[11,44] Please visit website[11] for more information.*

there are conditions where compelling information about effectiveness can be obtained without resort to RCTs.[23,30,42]

There are several types of RCTs.[11,43,44] CONSORT has developed a minimum set of "evidence-based" recommendations for evaluating, reporting and publishing trials; however, the recommendations "are not intended" for "designing, conducting, and analyzing trials."[11,44] The CONSORT flow diagram for the two group parallel RCT is shown in Figure 1, a type that now accounts for half of published trials,[44] an accompanying checklist (available on the website) helps readers and reviewers assess quality.[11]

Randomized controlled trials are "human constructs" and therefore "fallible;" hence, deficiencies in design, including errors because of cognitive biases, can occur at any phase of a trial,[30,45] the "poorly constructed trials" Sir Austin warned us about.[34] Despite the CONSORT recommendations, these errors can be missed by editors and reviewers of even "leading high impact journals,"[30] blind-siding the trusting clinician.

Methodological and statistical elements

Several methodological and statistical elements have become essential components of trials, RCTs included; they contribute to credible evidence when applied correctly. Conversely, limitations and violations can not only mislead when these are missed during the editorial and referee review process prior to publication, but also result in many RCTs being unhelpful.[30,46,47] Some of these are addressed by Sir Michael Rawlins.[23,31] Senn and Julious have discussed the several "unfortunate practices in measurements in clinical trials," from a statistical perspective.[48] Details are beyond the scope of this review. However, a summary of these issues for the clinician-reader may be informative.

1. The uncertainty principle, equipoise

The uncertainty principle and equipoise, although variably defined, reflect complementary ethical principles used to select subjects for trials, and have been described as 'moral underpinnings' of the RCT.[49-51] They require investigators, physician or other, to be ethically certain about the benefit-risk from investigating a particular treatment or test, and comparing it to either placebo or an existing established treatment or test. The recruiting investigator has not only to be well informed and convinced about the benefit-risk to her (his) patient or group of patients, but also must ensure that consents from potential participants and families are obtained ethically. Violations occurred more frequently in industry-sponsored trials than others in an older study,[52] but we are unaware of recent data.

2. Enrollment (assessment for eligibility)

Cognitive biases in the selection, consent and recruitment processes (even in RCTs) by investigators or their designates can skew results and influence generalizability. Biases are more likely if "allocation" is not concealed; selection biases may occur on the basis of literacy, gender, age, illness severity etc. Those with co-morbidity, women, children, the elderly and ethnic minorities are under-represented in most RCTs.[30] Biases in the obtaining of antenatal consent may skew results and affect

https://doi.org/10.1017/S0317167100014542 Published online by Cambridge University Press

generalizability of neonatal trials.[53] An example is discussed in the Appendix (#2) to Part 2.

3. The null hypothesis

The design and analyses of RCTs are commonly based on the null hypothesis: there is no difference in efficacy between the treatments.[23,31] The null hypothesis does not factor in adverse effects of the treatment under study. The use of the null hypothesis is questionable under the following situations which violate the hypothesis: (i) when there is previous evidence to suggest that the treatment under study is beneficial, and (ii) when trials are designed to show equivalence (no difference), no less benefit (non-inferiority), or not less than a pre-specified difference (futility).[23,31]

4. Sample size

The calculation of sample size is complex, the type of study being only one of several important determinants of methodology.[54-58] The objective is to recruit the 'right' number of subjects to get a meaningful clinical result. The concept of types I and II errors are familiar to most, but there is also a type III error that is not widely discussed.[55] These will be illustrated with examples in the appendices to Parts 1 (#1) and 2 (#2).

In a study conducted between 2005 and 2006, only 73 (34%) of 215 selected RCTs in "high impact" journals had complete data on sample size.[59] Despite the implicit assumption of accuracy, sample size may be "a guess masquerading as mathematics."[58] The current emphasis on sample size has been criticized, some authors calling for "more rational alternatives."[60,61]

5. Intention to treat (ITT) analysis

Sir Austin Bradford Hill referred to ITT, emphasizing the importance of complete follow up information.[62] All patients assigned to each of the treatment or placebo arms of a study are analyzed, including those who fail to complete the study or discontinue intervention; however, details for these drop outs must be provided (see Figure 1). Thus, ITT preserves randomization. However, once randomization occurs, ITT may not capture many events that can bias the study.[20] The lower the dropout rate, the more reliable the study, with 20% being arbitrarily considered the maximum. Violation, lack of clarity or inadequate description of ITT were found in 60%-75% of published RCTs analyzed in two studies; Hollis & Campbell cautioned readers to "critically assess validity of reported intention to treat analysis."[63-66]

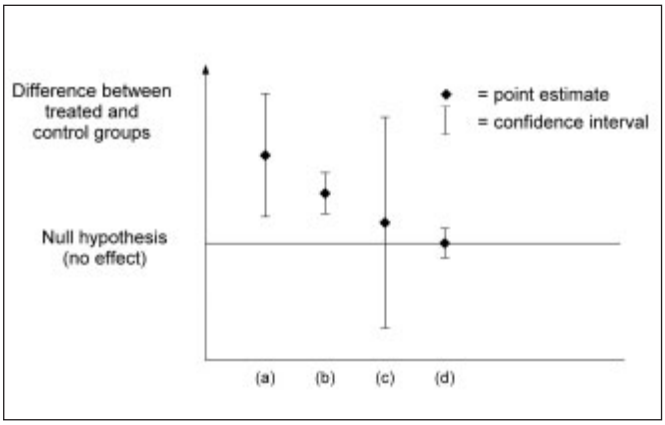

Figure 2: *Confidence intervals (CIs). (a) Statistically significant result (P < 0.05) but low precision, (b) statistically significant result (P < 0.05) with high precision, (c) not statistically significant result (P > 0.05) with low precision, and (d) not statistically significant result (P > 0.05) with high precision. Figure from Dr. Rafael Perera's presentation: "Making sense of results," at the spring 2012 Workshop on evidence-based practice, OCEBM. Reproduced with the kind permission of Dr. Perera and CEBM, Oxford, UK.*

6. Number needed to treat or harm (NNT or NNH)

The number needed to treat is the number of patients treated by the method under evaluation in a RCT, for one patient to benefit. Conversely, NNH is the number of patients similarly treated for one to be harmed. The NNT and NNH are specific to the study in question, for the characteristics of the subjects in the study, and for the time points in the course of treatment selected for their determination (Part 1 Appendix #2). Therefore, NNT and NNH from RCTs have to be extrapolated cautiously in clinical practice.[67] In addition, pooled numbers to treat from meta-analyses can be "seriously misleading" if there are differences in baseline risks and other characteristics between the included studies.[67,68] "Likelihood of being helped or harmed," "absolute and relative risk reduction or increase," are liable to similar limitations because of their statistical association with NNT and NNH. Reservations have been expressed about the methodological concepts underlying NNT.[48,69,70]

7. p-values

The p value gives the probability that a particular outcome would have arisen by chance, and is based on the null hypothesis. There are limitations to and errors in the use of the null hypothesis (see section on null hypothesis above), and in applying, reporting and interpreting p values.[23,71,72] Also, papers on RCTs often end with a statement about statistical significance without referring to clinical relevance, and a significant p value can bias readers about clinical importance.[73] Hence, we should be conscious of the truisms: (i) statistical significance is not synonymous with clinically important effects,[72] and (ii) significance usually pertains to group data not individual.

8. Confidence intervals (CIs)

The CI remains the optimal primary approach for assessing precision.[71,74,75] Confidence intervals quantifies uncertainty in or precision of measurement; the 95% CI is the range of values within which we can be 95% certain that the true value for the variable under study will lie (Figure 2). A wide CI should raise concern about the precision of data, and sample size (Part 1 Appendix #2 and #4). However, 'wide' is subjective.[76]

9. Surrogate markers

Surrogate markers are intended to substitute for a clinical outcome under study. Examples include glycated hemoglobin and blood sugars in diabetes mellitus and its complications, blood pressure and lipid values for cardiovascular disease and

https://doi.org/10.1017/S0317167100014542 Published online by Cambridge University Press

diabetes, and radiological findings for certain conditions.[77,78] Some neuroradiological surrogate markers include features of tumor, stroke, and multiple sclerosis; improvement or worsening on neuroradiological tests is often used to assess response to treatment in these conditions. When applied, interpreted and explained correctly, surrogate markers can reflect treatment response accurately; however, they may not always correlate with clinical outcome, including quality of life or side effects of drugs. Generally, they should not be used in isolation.[77,79-81] Surrogate markers like glycated hemoglobin or average of several blood sugar readings even in an individual diabetic, can miss hypoglycemia; surrogate markers failed to detect the adverse cardiovascular effects of glitazones.[77,82]

10. Composite outcomes

A composite outcome, such as death and hospital admission or death and 'major disability' (both ethically questionable in our view) is often used in RCTs. Several authors have drawn attention to important limitations to the use of composite outcomes.[83-88] The use of composite outcomes in trials can be problematic (Part 1 Appendix #3 and #4) because "outcomes are often unreasonably combined, inconsistently defined, and inadequately reported;" "readers may get an "exaggerated perception of how interventions work."[83]

11. Subgroup analysis

Subgroup analysis is an analysis of effects (example, of a treatment) on one or more subgroups of patients enrolled in a trial. Common subgroups include disease severity, age, gender, and ethnicity. In two studies, over two-thirds of published RCTs had deficiencies in analysis and reporting subgroup effects.[89,90] Subgroup analyses from a paper by Sacks et al[91] are briefly discussed in the appendix (Part 1 #4).

12. Stopping trials early

Randomized controlled trials may be stopped early, if benefit (as defined in the trial) is achieved or side effects occur. Beneficial effects can be significantly overestimated, and meta-analyses that include such trials will also be biased.[92-94] Guyatt et al provide examples and reasons, to illustrate how and why truncated trials can be seriously misleading; they suggest that any evidence from such trials should be classified as uncertain.[92] An example is discussed in the appendix to Part 2 (#2).

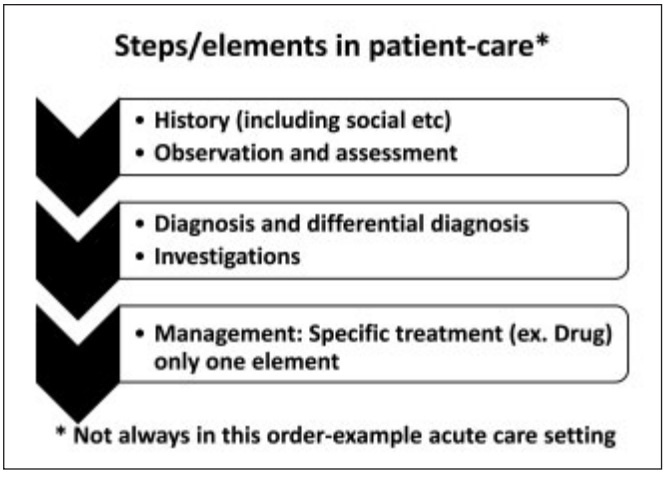

Figure 3: *Steps/elements in patient-care*.*

13. Post-hoc analysis

This refers to analyzing data after the study, and can help in formulating hypothesis, especially information about subgroups. They can only be used to draw tentative conclusions for confirmation with prospective studies, but not for making strong claims for effect. The paper by Cady et al,[95] illustrates the point (Part 1 Appendix #5).

Therapeutic devices and surgical procedures

There are additional issues that are unique to assessing the benefit and harm for therapeutic devices (neuromodulation is a fast growing field in the Neurosciences[96]) and surgical procedures.[38] For example, (i) design of RCTs has to be adapted to the surgical setting, (ii) valid information can be rendered out of date quickly, because of continuous improvements in technology, (ii) surgical experience and skill play a crucial role; for instance, there was apparently a 10 fold increase for in-hospital mortality in "real world" settings compared to trial settings for carotid endarterectomy, and (iii) most complications and mortality in surgical trials are perioperative, occurring early rather than late.[38] An example discussed in Part 2 (section on Neurosciences and Appendix #2 and #3) exemplifies the challenges of designing RCTs in neurosurgical critical care.

The challenges for Pediatrics (Neonatal to adolescent)

Many treatments are used "off-label" without direct evidence of efficacy or effectiveness in this age group. Neonates, children and adolescents are not little adults. A variety of developmental considerations, not only physiological and pharmacological, but also cognitive and psychological, must be factored in the design and conduct of trials.[97] In addition, the placebo response rate, at least in headache trials, is higher than in adults.[98] The mission of StaR Child Health, founded in 2009, is "to improve the design, conduct and reporting of pediatric research;"[99-101] the initiative is relevant for Pediatric clinical neurosciences. We hope StaR Child Health will incorporate the lessons learnt from two decades of the EBM paradigm.

Diagnosis and investigations

Treatment has been the dominant focus of EBM. However, history, observation, and examination are essential steps to diagnosis and investigations, before consideration of treatment (Figure 3).

Clinical diagnosis is complex and subjective.[2] Clinical decision-making (judgment) has intuitive and analytical elements, both being vulnerable to affective and other biases, and therefore susceptible to error.[102-106] The selection and interpretation of tests are also influenced by these factors. An observational study on how history can bias electrocardiograph interpretation in an emergency department is an excellent model for similar studies in the Neurosciences.[107,108] Awareness of our

https://doi.org/10.1017/S0317167100014542 Published online by Cambridge University Press

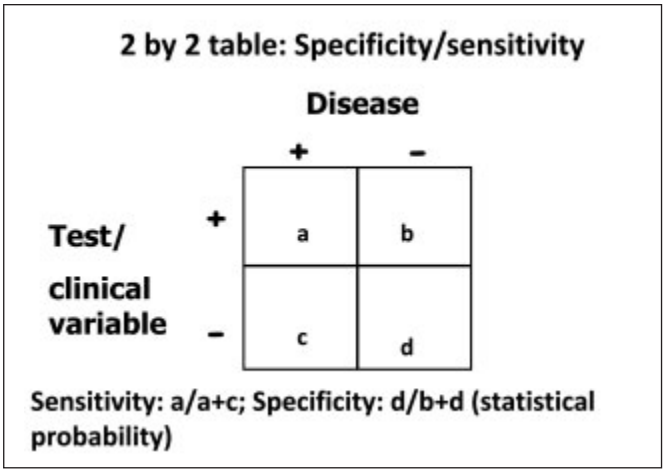

Figure 4: *2 by 2 table: Specificity/sensitivity. Sensitivity: proportion of subjects with the disease who have a positive test; Specificity: proportion of subjects without the disease who have a negative test; Positive predictive value (PPV): proportion of subjects who test positive and actually have the disease: a/a+b; Negative predictive value (NPV): proportion of subjects who test negative and don't have the disease: d/c+d*

biases may minimize diagnostic errors. We need further insights into clinical decision-making, so that errors can be minimized.[102,103]

The calculation of specificity, sensitivity, likelihood and related ratios, p values, and CIs to assess and apply diagnostic tests to populations and patients, are other cornerstones of the EBM paradigm.[2,109] There are limitations to such methods:[110] (i) Typically, the calculations are done with a "training" set for whom the diagnosis is known, but then applied to an independent set or individual subjects in whom the diagnosis is uncertain, rendering the predictive value unreliable; (ii) tests of specificity and sensitivity, are usually dichotomous i.e., positive or negative; present or absent (Figure 4); there are many grey zones in clinical practice; (iii) a test may be negative for the disease under study, but may have revealed diagnostic information about another, or been used to exclude an important differential diagnosis, (iv) the quality of the reference standard used for comparison may be questionable, (v) many tests, especially electrophysiological, pathological, and radiological, are not only dependent on human interpretation, but also have grey areas; most importantly, in clinical practice tests are not used in isolation, but combined with history, physical examination and other tests to draw conclusions.

Systematic reviews-SRs (including Meta-analyses) and guidelines

Meta-analyses refer to combining (data from) studies; such a review can also be termed systematic quantitative review.[30] When data from several studies are analyzed and discussed without pooling them, the review is termed systematic qualitative review.[30] A SR represents the consolidated evidence on a particular question, but the quality of the evidence can only reflect the quality of the incorporated studies; therefore, SRs may be flawed if they include trials that are deficient or biased.[15,30,68,93,94,111-114] Additionally, Shojania et al found that 57% of 100 high quality SRs reviewed had out-dated information, and suggested that developers of SRs ensure timely revisions.[115] Classification tools are often used to assess the design of studies for inclusion in SRs; one study found moderate inter-rater reliability and low accuracy for one such tool.[116] If generalizable, this can create another bias in some SRs. Hence, although considered the gold standard for evidence in some classifications, SRs "are not infallible."[38] A SR is discussed in Part 2 (Appendix #1).

The caveats, "garbage in garbage out" and timely updating, also apply to clinical practice guidelines. In addition, guidelines may not apply to populations (such as children, the elderly, female or specific ethnic groups) excluded from or to issues (for instance, co-morbidities) not specifically examined in the studies on which they are based.[117-119] Guidelines are not meant to be prescriptive.[8]

Cochrane reviews

Because of their promise,[9] great faith is placed in Cochrane reviews, justifiably in many situations. However, Cochrane reviews are susceptible to the same potential biases as other SRs.[20] Some have been challenged for their accuracy.[113,120,121] Other issues include: (1) variability in how abstracts are written;[122] (2) apparently, it is not mandatory for authors to report conflicts of interest in the trials included for review,[123,124] (3) reviews can be out-dated,[30,115,122] (4) the Risk of Bias (ROB) tool is meant to detect bias in studies being considered for inclusion; inter-rater reliability for this tool was low for almost all domains, potentially contributing to flaws,[125] and (5) to our knowledge, Cochrane Collaboration does not provide explicit information about reviews that have been withdrawn because of potentially misleading information (Cochrane Collaboration; personal communication); many journals do.

Conflicts of interest, ethics

Conflicts of interest and ethical violations compromise trustworthiness of information from trials, and can put patients at risk. These have been addressed by Tharyan; some researchers have been guilty of fraud, but there are other issues also.[30,126,127] Montori and Guyatt drew attention to the "corruptions in the evidence base", including (i) the misuse of the terms and concepts of EBM in guidelines and algorithms, (ii) industry funded investigators, interpreting research in favor of industry, and (iii) misleading descriptions or presentations of research findings and violations of several components involved in the research process.[5]

Additionally, there is "honorary authorship" (i.e. name of author included when he/she has not made a contribution to the paper; often to add prestige), and "ghost authorship" (authors who have made significant contributions to the paper, are not included, often those writing for industry sponsored trials); many of these papers have been published in high impact journals.[126-130] Honorary and ghost authorships have occurred with Cochrane reviews,[128-130] but do not involve industry related conflict of interest. These data suggest that current "conflict of interest" declarations are inadequate.[128-130]

Conflicts of interest in guidelines,[131-135] and meta-analyses,[123] are also being reported. Similar concerns have been voiced in the

https://doi.org/10.1017/S0317167100014542 Published online by Cambridge University Press

surgical literature,[136] and journals and publishers also have potential conflicts of interest.[137,138]

There are serious consequences to patient care from undisclosed company data and from the biases in industry funded trials and publications. Reported examples include those related to rifocoxib, rosiglitazone, oseltamivir, and gabapentin, involving four separate leading pharmaceutical companies.[82,139-142]

The corporate sector does not have a monopoly on greed, and non-profit organizations can be just as venal.[143] Booth has compared the marketing strategies of some EBM groups to that of "the manufacturer of the world's most famous soft drink."[144] Some examples to support his opinion are provided in the appendix (Part 1 #6).

Conclusions

In this the first of two parts, we have discussed the knowledge gained over the first twenty years of the EBM paradigm. Related issues and suggestions are discussed in Part 2.

References

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268:2420-5.
2. Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-based Medicine; 4th ed. Toronto: Churchill Livingstone Elsevier, 2011.
3. Goldenberg MJ, Borgerson K, Bluhm R. The nature of evidence in evidence-based medicine: guest editors' introduction. Perspect Biol Med. 2009;52:164-7.
4. Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic inquiries in evidence-based medicine. Cancer Control. 2009;16:158-68.
5. Montori VM, Guyatt GH. Progress in evidence-based medicine. JAMA. 2008;300:1814-6.
6. Burneo JG, Demaerschalk BM, Jenkins ME, editors. Neurology: An evidence-based approach. New York: Springer; 2012.
7. Wiebe S, Demaerschalk B. Evidence based care in the neurosciences. Can J Neurol Sci. 2002;29:115-9.
8. 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.
9. Cochrane Collaboration. [Cited 2012 Aug 27]. Available from: http://www.cochrane.org/cochrane-reviews/
10. Centre for Evidence-Based Medicine, University of Oxford. Oxford, UK. [Cited 2012 Dec 12]. Available from: http://www.cebm.net/
11. CONSORT group. [Cited 2012 Dec 11]. Avaiable from: http://www.consort-statement.org/
12. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. BMJ. 1995;310:1085-6.
13. 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.
14. Straus SE, McAlister FA. Evidence-based medicine: a commentary on common criticisms. CMAJ. 2000;163:837-41.
15. Elamin MB, Montori VM. The hierarchy of evidence: From unsystematic clinical observations to systematic reviews. Chapter 2. In: Burneo JG, Demaerschalk BM, Jenkins ME, editors. Neurology: An evidence-based approach. New York: Springer; 2012:p.11-24.
16. Anonymous. The periodic health examination: 2. 1985 update. Canadian Task Force on the Periodic Examination. CMAJ. 1986;134:724-7.
17. French J, Gronseth G. Lost in a jungle of evidence: we need a compass. Neurology. 2008;71:1634-8.
18. Lee SK, Singhal N, Aziz K, Cronin CM. The EPIQ evidence reviews - practical tools for an integrated approach to knowledge translation. Paediatr Child Health. 2011;16:629-30.
19. 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.
20. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". Am J Med. 1997;103:529-35.
21. Bluhm R, Borgerson K. Evidence-based Medicine. In: Gifford F, editor. Handbook of the Philosophy of Science. Volume 16: Philosophy of Medicine. Amsterdam: Elsevier, BV; 2011: p. 203-38.
22. Cohen AM, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of Evidence-Based Medicine. Int J Med Inform. 2004;73:35-43.
23. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. Lancet. 2008;372:2152-61.
24. Rosner AL. Evidence-based medicine: revisiting the pyramid of priorities. J Bodyw Mov Ther. 2012;16:42-9.
25. Tonelli MR. The limits of evidence-based medicine. Respir Care. 2001;46:1435,40; discussion 1440-1.
26. Upshur RE. Are all evidence-based practices alike? Problems in the ranking of evidence. CMAJ. 2003;169:672-3.
27. Upshur R, Tracy CS. Evidence-based medicine: perils and pitfalls. N Z Fam Physician. 2003;30(5):327-30.
28. Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? NeuroRx. 2004;1:341-7.
29. Goodin DS, Reder AT. Evidence-based medicine: promise and pitfalls. Mult Scler. 2012;18:947-8.
30. Jadad AR, Enkin MW. Randomized controlled trials. Oxford, UK: Blackwell Publishing BMJI Books; 2007.
31. Rawlins MD. De Testimonio: On the evidence for decisions about the use of therapeutic interventions. London, UK: Royal College of Physicians; 2008.
32. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490.
33. 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.
34. Hill AB. Reflections on controlled trial. Ann Rheum Dis. 1966;25:107-13.
35. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis. 1967;20:637-48.
36. Cochrane AL. Archie Cochrane in his own words. Selections arranged from his 1972 introduction to "Effectiveness and Efficiency: Random Reflections on the Health Services" 1972. Control Clin Trials. 1989;10:428-33.
37. Horwitz RI. The dark side of evidence-based medicine. Cleve Clin J Med. 1996;63:320-3.
38. Hartling L, McAlister FA, Rowe BH, Ezekowitz J, Friesen C, Klassen TP. Challenges in systematic reviews of therapeutic devices and procedures. Ann Intern Med. 2005;142:1100-11.
39. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci. 2011;13:217-24.
40. Buchanan WW, Kean WF. Evidence based medicine: the median is not the message. J Rheumatol. 2001;28:2371-2.
41. Brigo F. New anti-epileptic drugs: overcoming the limits of randomised controlled trials. Int J Evid Based Healthc. 2011;9:440-3.
42. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ. 2007;334:349-51.
43. Wikipedia. The randomized controlled trial. [Cited 2012 Dec 12]. Available at: http://en.wikipedia.org/wiki/Randomized_controlled_trial
44. 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.
45. Makhinson M. Biases in the evaluation of psychiatric clinical evidence. J Nerv Ment Dis. 2012;200:76-82.
46. Zwarenstein M, Oxman A. Pragmatic Trials in Health Care Systems (PRACTIHC). Why are so few randomized trials useful, and what can we do about it? J Clin Epidemiol. 2006;59:1125-6.
47. Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. Arch Intern Med. 2008;168:133-5.

48. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? Stat Med. 2009;28:3189-209.
49. Gifford F. Uncertainty about clinical equipoise. Clinical equipoise and the uncertainty principles both require further scrutiny. BMJ. 2001;322:795.
50. Oricha BS, Yauri MB. Uncertainty principle versus clincal equipoise in clinical trials in Sub-Saharan Africa: Are they really tenable? Ann African Med. 2003;2(2):99-100.
51. Weijer C, Shapiro SH, Cranley Glass K. For and against: clinical equipoise and not the uncertainty principle is the moral underpinning of the randomised controlled trial. BMJ. 2000;321: 756-8.
52. Djulbegovic B, Lacevic M, Cantor A, et al. The uncertainty principle and industry-sponsored research. Lancet. 2000;356: 635-8.
53. Rich W, Finer NN, Gantz MG, et al. Enrollment of extremely low birth weight infants in a clinical research study may not be representative. Pediatrics. 2012;129:480-4.
54. Biau DJ, Kerneis S, Porcher R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clin Orthop Relat Res. 2008;466:2282-8.
55. Heinzl H, Benner A, Ittrich C, Mittlbock M. Proposals for sample size calculation programs. Methods Inf Med. 2007;46:655-61.
56. van der Tweel I, Askie L, Vandermeer B, et al. Standard 4: determining adequate sample sizes. Pediatrics. 2012;129 Suppl 3:S138-45.
57. Sakpal TV. Sample size estimation in clinical trial. Perspect Clin Res. 2010;1:67-9.
58. Julious SA, Campbell MJ. Tutorial in biostatistics: sample sizes for parallel group clinical trials with binary data. Stat Med. 2012;31: 2904-36.
59. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ. 2009;338:b1732.
60. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Med. 2010;8:17.
61. Bacchetti P, Deeks SG, McCune JM. Breaking free of sample size dogma to perform innovative translational research. Sci Transl Med. 2011;3:87ps24.
62. Hill AB. Medical ethics and controlled trials. BMJ. 1963;1: 1043-9.
63. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ. 1999; 319:670-4.
64. Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? Clin Trials. 2007;4:350-6.
65. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ. 2010;340: c2697.
66. Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: Who is in? Who is out? J Fam Pract. 2002;51:969-71.
67. McAlister FA. The "number needed to treat" turns 20--and continues to be used and misused. CMAJ. 2008;179:549-53.
68. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses--sometimes informative, usually misleading. BMJ. 1999;318:1548-51.
69. Hutton JL. Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. Br J Haematol. 2009;146:27-30.
70. Julious SA. Issues with number needed to treat. Stat Med. 2005;24: 3233-5.
71. Greenhalgh T. How to read a paper. Statistics for the non-statistician. II: "Significant" relations and their pitfalls. BMJ. 1997;315:422-5.
72. Goodman S. A dirty dozen: twelve p-value misconceptions. Semin Hematol. 2008;45:135-40.
73. Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values on the perceived importance of study results. Acta Orthop. 2005;76:291-5.
74. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. CMAJ. 1995;152:169-73.
75. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011;64:1283-93.
76. Glasziou P, Doll H. Was the study big enough? Two cafe rules. Evid Based Med. 2006;11:69-70.
77. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011;343:d7995.
78. Aronson JK. Surrogate end points: studying their benefits, taxonomy, and semantics. BMJ. 2012;344:e750.
79. Colatsky TJ. Reassessing the validity of surrogate markers of drug efficacy in the treatment of coronary artery disease. Curr Opin Investig Drugs. 2009;10:239-44.
80. Psaty BM, Weiss NS, Furberg CD, et al. Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. JAMA. 1999;282: 786-90.
81. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125:605-13.
82. Montori VM, Shah ND. What have we learnt from the rosiglitazone saga? BMJ. 2011;342:d1354.
83. Cordoba G, Schwartz L, Woloshin S, Bae H, Gotzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. BMJ. 2010;341:c3920.
84. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007; 334:786.
85. Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. J Clin Epidemiol. 2007;60:651,7; discussion 658-62.
86. Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, et al. Composite outcomes can distort the nature and magnitude of treatment benefits in clinical trials. Ann Intern Med. 2009;150: 566-7.
87. Montori VM, Busse JW, Permanyer-Miralda G, Ferreira I, Guyatt GH. How should clinicians interpret results reflecting the effect of an intervention on composite endpoints: should I dump this lump? ACP J Club. 2005;143:A8.
88. Montori VM, Permanyer-Miralda G, Ferreira-Gonzalez I, et al. Validity of composite end points in clinical trials. BMJ. 2005; 330:594-6.
89. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 2012;344:e1553.
90. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med. 2007;357:2189-94.
91. Sacks FM, Pfeffer MA, Moye LA, et al. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. N Engl J Med. 1996;335:1001-9.
92. Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM. Problems of stopping trials early. BMJ. 2012;344:e3863.
93. Bassler D, Ferreira-Gonzalez I, Briel M, et al. Systematic reviewers neglect bias that results from trials stopped early for benefit. J Clin Epidemiol. 2007;60:869-73.
94. Bassler D, Montori VM, Briel M, et al. Reflections on meta-analyses involving trials stopped early for benefit: Is there a problem and if so, what is it? Stat Methods Med Res. 2013;22: 159-68.
95. Cady RK, Lipton RB, Hall C, Stewart WF, O'Quinn S, Gutterman D. Treatment of mild headache in disabled migraine sufferers: results of the Spectrum Study. Headache. 2000;40:792-7.
96. Tellez-Zenteno JF, Wiebe S. Hippocampal stimulation in the treatment of epilepsy. Neurosurg Clin N Am. 2011;22:465,75, vi.
97. Andrasik F, Powers SW, McGrath PJ. Methodological considerations in research with special populations: children and adolescents. Headache. 2005;45:520-5.
98. Lewis DW, Winner P, Wasiewski W. The placebo responder rate in children and adolescents. Headache. 2005;45:232-9.

99. Klassen TP, Hartling L, Craig JC, Offringa M. Children are not just small adults: the urgent need for high-quality trial evidence in children. PLoS Med. 2008;5:e172.
100. Klassen TP, Hartling L, Hamm M, van der Lee JH, Ursum J, Offringa M. StaR Child Health: an initiative for RCTs in children. Lancet. 2009;374:1310-2.
101. Hartling L, Wittmeier KD, Caldwell PH, et al. StaR Child Health: developing evidence-based guidance for the design, conduct, and reporting of pediatric trials. Clin Pharmacol Ther. 2011;90: 727-31.
102. Croskerry P. Perspectives on diagnostic failure and patient safety. Healthc Q. 2012;15 Suppl:50-6.
103. Croskerry P. A universal model of diagnostic reasoning. Acad Med. 2009;84:1022-8.
104. Croskerry P. Context is everything or how could I have been that stupid? Healthc Q. 2009;12 Spec No Patient:e171-6.
105. Croskerry P, Abbass AA, Wu AW. How doctors feel: affective issues in patients' safety. Lancet. 2008;372:1205-6.
106. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. BMJ. 2009;338:b946.
107. Cruz MF, Edwards J, Dinh MM, Barnes EH. The effect of clinical history on accuracy of electrocardiograph interpretation among doctors working in emergency departments. Med J Aust. 2012; 197:161-5.
108. Katelaris A. What influences clinical decision making? Med J Aust. 2012;197:129.
109. Perera R, Heneghan C, Badenoch D. Statistics toolkit. Oxford, UK: Blackwell publishing; 2008.
110. Feinstein AR. Misguided efforts and future challenges for research on "diagnostic tests". J Epidemiol Community Health. 2002;56: 330-2.
111. Stegenga J. Is meta-analysis the platinum standard of evidence? Stud Hist Philos Biol Biomed Sci. 2011;42:497-507.
112. Coyne JC, Thombs BD, Hagedoorn M. Ain't necessarily so: review and critique of recent meta-analyses of behavioral medicine interventions in health psychology. Health Psychol. 2010;29: 107-16.
113. Clark DA. The end of evidence-based medicine? Inflammopharmacology. 2012;20:187-93.
114. 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.
115. Shojania KG, Sampson M, Ansari MT, et al. Technical review number 16: Updating systematic reviews. AHRQ Publication No. 07-0087. Rockville (MD): U.S. Department of Health and Human Services; 2007.
116. Hartling L, Bond K, Santaguida PL, Viswanathan M, Dryden DM. Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy. J Clin Epidemiol. 2011;64:861-71.
117. Mutasingwa DR, Ge H, Upshur RE. How applicable are clinical practice guidelines to elderly patients with comorbidities? Can Fam Physician. 2011;57:e253-62.
118. Lugtenberg M, Burgers JS, Clancy C, Westert GP, Schneider EC. Current guidelines have limited applicability to patients with comorbid conditions: a systematic analysis of evidence-based guidelines. PLoS One. 2011;6:e25987.
119. Ballantyne AJ, Rogers WA. Sex bias in studies selected for clinical guidelines. J Womens Health (Larchmt). 2011;20:1297-306.
120. Fahy K, Tracy SK. Critique of Cochrane systematic review of home-like setting for birth. Int J Evid Based Healthc. 2007; 5:360-4.
121. Cundiff DK. A systematic review of Cochrane anticoagulation reviews. Medscape J Med. 2009;11:5.
122. Bow S, Klassen J, Chisholm A, et al. A descriptive analysis of child-relevant systematic reviews in the Cochrane Database of Systematic Reviews. BMC Pediatr. 2010;10:34.
123. Roseman M, Milette K, Bero LA, et al. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. JAMA. 2011;305:1008-17.
124. Roseman M, Turner EH, Lexchin J, Coyne JC, Bero LA, Thombs BD. Reporting of conflicts of interest from drug trials in Cochrane reviews: cross sectional study. BMJ. 2012;345:e5155.
125. 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.
126. Tharyan P. Evidence-based medicine: can the evidence be trusted? Indian J Med Ethics. 2011;8:201-7.
127. Tharyan P. Criminals in the citadel and deceit all along the watchtower: Irresponsibility, fraud, and complicity in the search for scientific truth. Mens Sana Monogr. 2012;10:158-80.
128. Bonita RE, Adams S, Whellan DJ. Reporting of clinical trials: publication, authorship, and trial registration. Heart Fail Clin. 2011;7:561-7.
129. Mowatt G, Shirran L, Grimshaw JM, et al. Prevalence of honorary and ghost authorship in Cochrane reviews. JAMA. 2002;287: 2769-71.
130. Wislar JS, Flanagin A, Fontanarosa PB, Deangelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ. 2011;343:d6128.
131. Avraham R. Clinical practice guidelines: the warped incentives in the U.S. healthcare system. Am J Law Med. 2011;37:7-40.
132. Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010;152:738-41.
133. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011;343:d5621.
134. Scott IA, Guyatt GH. Clinical practice guidelines: the need for greater transparency in formulating recommendations. Med J Aust. 2011;195:29-33.
135. Norris SL, Holmer HK, Burda BU, Ogden LA, Fu R. Conflict of interest policies for organizations producing a large number of clinical practice guidelines. PLoS One. 2012;7:e37413.
136. Bailey CS, Fehlings MG, Rampersaud YR, Hall H, Wai EK, Fisher CG. Industry and evidence-based medicine: Believable or conflicted? A systematic review of the surgical literature. Can J Surg. 2011;54:321-6.
137. Lundh A, Barbateskovic M, Hrobjartsson A, Gotzsche PC. Conflicts of interest at medical journals: the influence of industry-supported randomised trials on journal impact factors and revenue - cohort study. PLoS Med. 2010;7:e1000354.
138. Upshur R, Buetow S, Loughlin M, Miles A. Can academic and clinical journals be in financial conflict of interest situations? The case of evidence-based incorporated. J Eval Clin Pract. 2006;12:405-9.
139. Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations from the Tamiflu experience. PLoS Med. 2012;9:e1001201.
140. Moynihan R. Rosiglitazone, marketing, and medical science. BMJ. 2010;340:c1848.
141. Wang AT, McCoy CP, Murad MH, Montori VM. Association between industry affiliation and position on cardiovascular risk with rosiglitazone: cross sectional systematic review. BMJ. 2010;340:c1344.
142. Zarin DA, Tse T. Medicine. Moving toward transparency of clinical trials. Science. 2008;319:1340-2.
143. Rhode DL, Packel AK. Ethics and Non Profits. [updated summer 2009; cited 2012 Aug 5] Available at: http://www.ssireview.org/articles/entry/ethics_and_nonprofits
144. Booth A. Evidence-based practice: triumph of style over substance? Health Info Libr J. 2011;28:237-41.

Appendix to Part 1

Some examples are included here to complement the text but avoid cluttering it. Readers will need to have the full length manuscripts of the examples discussed below, to fully understand the context, and make their own critical appraisal. Ours are open to critique. The numbering of references corresponds to the first citation in the text.

1. Sample size

Let us compare two therapies, A and B: A and B are assumed to be equally effective in terms of benefit (The null hypothesis: A=B in efficacy; note that side-effects are not factored in).

(i) Type I (false positive; α) error refers to a situation when a statistical difference (A better than B or B better than A) is erroneously found, when there is actually none; the lower the level of significance, the lower the probability of type I error; conventionally, the level of significance is set at 5%. The level could be set lower (example 1%) if considered clinically necessary, as when a new treatment with unknown side-effects is compared to an established effective treatment with low side-effects. Lower levels requiring a larger sample size.

(ii) Type II (false negative; β) error is one when no difference is found, although one actually exists i.e., in reality A is better than B or B is better than A; the assumption of the study is that A and B are equivalent because there is no information to the contrary. The power refers to the chance (10% if the power is set at 90%; by convention, power set at 80%) the study will not detect the difference. The greater the power the greater the probability of detecting the difference, and larger the sample size needed.

(iii) Type III error. Despite the assumption of no difference between A and B, A is actually better than B (i.e., B is inferior to A). In the analysis, B (the inferior drug) is found to be statistically superior to A (the better drug); an underpowered study contributes to type III error.[55]

2. Number needed to treat (NNT)

NNT is calculated for a specific period in the course of a trial: In a study on statins, the NNT to prevent one coronary event was 364 (95% CI 210-1362) for those treated for 90 days, and 93 (95% CI 59-208) for those treated for 3.3 years.[67] As an aside, this example also suggests imprecision of data because of the wide confidence intervals (CI; see discussion on CI in the text, Figure 2 and examples that follow).

3. Composite outcomes

A drug leads to a large reduction in the composite outcome of death or chest pain. This could mean the treatment resulted in both fewer deaths and chest pain, either death or chest pain reduced or unchanged but the other increased (so death could be increased but chest pains decreased in the group as a whole).[83]

4. Subgroup analyses

Sacks et al reported on a placebo-controlled statin trial,[91] a paper that provides an excellent opportunity to address several issues that have been discussed in our review, including subgroup analyses. Please refer to Tables 1, 2 and 3 in their paper.

(i) The primary outcome was a composite of death from coronary heart disease and non fatal myocardial infarction, a combination that can be questioned both ethically and clinically. Despite statistical significance, the 95% CI was wide at 9-36. There was no difference in deaths and the CI was wide at -5 to 39. The CIs were wide for all the other ten prespecified outcome subgroups.

(ii) Sixteen subgroup analyses were done (Table 3 in their paper) for base-line variables including gender, age, presence or absence of hypertension etc. Only 14% were female and only 7.5% were non whites. The CIs for all the subgroups were wide.

(iii) Lipid levels included in the subgroup analyses may be considered a surrogate marker for a clinical outcome.

The rather wide CIs for all values, some more than others, would suggest imprecision of data (Figure 2), and in our opinion, should have warranted a more cautious conclusion than the one the authors came to from this study. In addition, attention should have been drawn to the small number of females and non-whites, to avoid unwarranted extrapolation to these populations.

5. Post-hoc analysis

Cady et al[95] published a post-hoc analysis of a subgroup of 26 patients from a larger study. In this subgroup with disabling headache (International headache society criteria for migraine codes 1.1 and 1.2), the authors found a trend for better response and less recurrence when mild headaches were treated than when moderate or severe headaches were treated; they speculated that in those with disabling migraine, early treatment would be more beneficial than later treatment. The authors (correctly) referred to the post-hoc analysis and small sample size in discussion, and suggested that larger RCTs were needed to confirm their opinion.

6. Statements that may support Booth's reference to marketing comparable to that of a leading multi-national soft drinks manufacturer.[144]

(A) To cite from Straus et al: EBM "has grown exponentially,... and evidence-based practice has been incorporated into most if not all healthcare disciplines...adding evidence-based to the title of a book can increase sales...there has been an explosion in the number of courses, workshops, and seminars...EBM educational interventions for the public, policy makers and healthcare managers have grown."[2]

(B) From the website of the Cochrane Collaboration[9]: (i) "Every day someone, somewhere searches the Cochrane library every two seconds, reads an abstract every two seconds...," (ii) "No other organisation matches the quality...Cochrane Reviews," (iii) "Without Cochrane reviews, people making decisions are unlikely to be able to access and make full use of existing healthcare research," (iv) "Cochrane reviews enable the practice of evidence-based health care," and (v) "Users of the medical literature should start paying more attention to the Cochrane database of systematic reviews.., and less attention to some better known competitors."

Comment: The impression given is that the products are infallible, when they are not (see text in both Parts 1 and 2). The statements may be taken at face value by the trusting reader including the public and policy makers, who may not be able to evaluate the evidence for himself or herself.

https://doi.org/10.1017/S0317167100014542 Published online by Cambridge University Press