Journal of Bodywork & Movement Therapies (2012) 16, 42–49

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/jbmt

CRITICAL REVIEW

Evidence-based medicine: Revisiting the pyramid of priorities

Anthony L. Rosner, Ph.D., LL.D.[Hon.], LLC*

International College of Applied Kinesiology, USA

Received 5 February 2011; received in revised form 7 May 2011; accepted 10 May 2011

KEYWORDS
Clinical thinking;
Scientific medicine;
Research evidence;
Evidence informed medicine

Summary Evidence-based medicine (EBM) is beset with numerous problems. In addition to the fact that varied audiences have each customarily sought differing types of evidence, EBM traditionally incorporated a hierarchy of clinical research designs, placing systematic reviews and meta-analyses at the pinnacle. Yet the canonical pyramid of EBM excludes numerous sources of research information, such as basic research, epidemiology, and health services research. Models of EBM commonly used by third party payers have ignored clinical judgment and patient values and expectations, which together form a tripartite and more realistic guideline to effective clinical care. Added to this is the problem in which enhanced placebo treatments in experimentation may obscure verum effects seen commonly in practice. Compounding the issue is that poor systematic reviews which comprise a significant portion of EBM are prone to subjective bias in their inclusion criteria and methodological scoring, shown to skew outcomes. Finally, the blinding concept of randomized controlled trials is particularly problematic in applications of physical medicine. Examples from the research literature in physical medicine highlight conclusions which are open to debate. More progressive components of EBM are recommended, together with greater recognition of the varying audiences employing EBM.
© 2011 Elsevier Ltd. All rights reserved.

Definitions of EBM

"Evidence-based medicine" (EBM) was introduced as a term to denote the application of treatment that has been proven and tested "in a rigorous manner to the point of its becoming 'state of the art'" (Gupta, 2003). Its intention has been to ensure that the information upon which doctors and patients make their choices is of the highest possible standard (Coulter et al., 1999). To reach a clinical decision based upon the soundest scientific principles, EBM proposes five steps for the clinician to follow as shown in Table 1 (Sackett et al., 2000). *Step 2* (accessing the best evidence) customarily follows a totemic relationship of the available

* 1330 Beacon Street, Suite #315, Brookline, MA 02446-3202, USA. Tel.: +1 617 734 3397; fax: +1 617 734 0989.
E-mail address: arosner66@aol.com.

1360-8592/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jbmt.2011.05.003

Table 1 Steps of evidence-based medicine (Sackett et al., 2000).

Step	Process
1.	Converting the need for information [about prevention, diagnosis, prognosis, therapy, causation, etc.] into an answerable question [Also known as formulating a clinical question]
2.	Finding the best evidence with which to answer that question
3.	Critical appraisal of the research evidence for validity, impact, and application
4.	Integration of the critically appraised research evidence with the expertise of the practitioner and the patient's unique biology, values and circumstances
5.	Evaluating the effectiveness of the process in improving patient care and practitioner efficiency in executing steps #1—4 and seeking ways to improve both

designs of clinical research, portrayed as a pyramid as shown in Figure 1a (http://phpartners.org/tutorial/04-ebph2/2-keyConcepts/imagines/4.2.7.1.jpg).

Here it is evident that systematic reviews and meta-analyses occupy the top echelon in predicting clinical outcomes, followed by randomized controlled double-blind studies (RCTs) and thence by cohort studies, case control studies, case series, and case reports. The fact that not all these levels are taken into consideration in certain renderings of EBM is problematic. Particularly bedeviling to RCTs is that enhanced placebo effects sometimes obscure those of the verum interventions. Furthermore, the "usefulness" of information gleaned from different experimental designs is inextricably bound to the specific purposes to which they are applied. Finally, such elements as clinical judgment and patient values, together with research reports forming a three-legged stool of elements comprising a more progressive construct of EBM, have been neglected. All these concepts will be examined below.

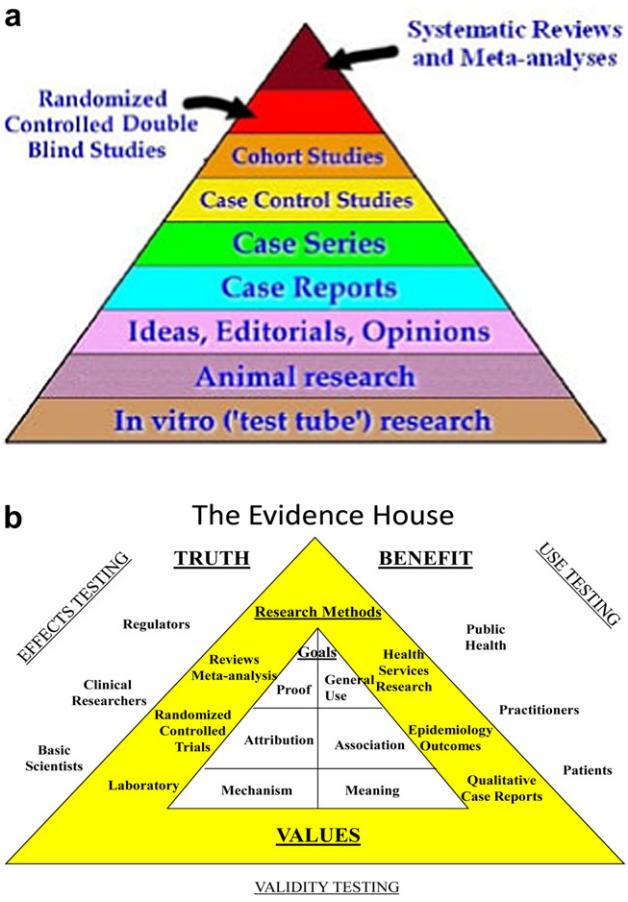

Figure 1 a: The hierarchy of clinical evidence. b: The evidence house (Jonas, 2001).

Traditional building blocks of EBM pyramid

Systematic reviews and meta-analyses, as opposed to simple narrative reviews, not only attempt to screen the largest sample of reports of potential interest, but also vet them for their rigor and select only the articles presumed to be the most valid pertaining to a focused research question. Particularly with meta-analyses, such reviews attempt to reconcile contradictory reports by combining specific outcome measures statistically in the attempt to reach a final level of significance of treatment effects.

Randomized controlled trials (RCTs) draw from a target population presumed to represent typical patients affected by the given condition under study, dividing the resulting sample randomly into two or more groups receiving specific treatments. One such treatment may be a placebo or sham intervention, while one or more other interventions involve specific modalities or treatment packages to which the experimenter seeks to ascribe specific effects from the outcomes obtained. Central to the design of the traditional RCT for over half a century is the concept of blinding of patient and practitioner, in which the assignment and dispensation of treatment regimens remains masked (Aravanis and Luisada, 1956). The advantage of randomization is to equalize as many effects as possible that do not involve the intervention across all groups compared, thereby minimizing unanticipated elements that would confound the results obtained from the treatments themselves. It is important to remember that RCTs evaluate average effects, determining intervention effects upon groups of subjects rather than the individual.

Case reports and case series present what actually has occurred in clinical practice as narrative reports of the presentation, treatment, and progression of either an individual or groups of patients. Unlike RCTs, case reports emphasize the individual rather than the group. The purpose of case studies is not only to present materials to other practitioners relating to patient management, but also to uncover areas requiring further exploration. While incapable of testing hypotheses, they offer complete

narratives of real-life clinical situations which may generate hypotheses for further research.

Limitations of EBM pyramid and evolution to more progressive "evidence informed" models

Questions about the pecking order of certain types of clinical studies predicting clinical outcomes as ranked in the EBM pyramid (Figure 1a) began to appear in the 1980s, when the quality of lowly observational (cohort, case series) studies was found to improve, such that their predictive value in clinical situations could now be compared to that seen in the more rigorous RCTs (Benson and Hartz, 2000; Concato et al., 2000). At the same time, lofty RCTs began to be seriously challenged due to their limited applicability in clinical situations (Wallach et al., 2002; Tonelli, 1998).

In other words, patients studied in RCTs are often not the patients seen in everyday practice. Essentially, efficacy rather than effectiveness is more commonly emphasized, as the generalizability of RCTs is limited to study groups of varying homogeneities. For example, patients with comorbidities are often excluded in RCTs in order to obtain a homogeneous sample, and it is not at all clear that the patient who even agrees to participation in an RCT is truly representative of the general population (Wallach et al., 2002). Particularly vexing to RCTs is the observation that enhanced placebo effects may constitute most of the verum effect, such that what appears to be a highly effective treatment in practice can be masked when compared to a placebo intervention in experimentation, previously described as "double positive paradox" (Walach, 2001; Reilly, 2002). Lack of insight into lifestyles, nutritional interventions, and long-latency deficiency diseases (Heany, 2003) have also plagued RCTs. Ignored lifestyle choices, for example, have triggered the proliferation of chronic diseases (Pronk et al., 2010). Serious flaws have likewise surfaced which demonstrate how even the exalted meta-analysis is subject to human error and bias (Juni et al., 1999; Rosner, 2003). Finally, the double-blind protocol established as a core to the RCT for over half a century in allopathic medicine in the study of pharmaceuticals or surgery (Aravanis and Luisada, 1956) becomes a virtual impossibility in applications of physical medicine.

Adding to the traditional elements of evidence as shown in the traditional evidence pyramid has been the pronouncement of the epidemiologist David Sackett, who in 1997 introduced the *clinical judgment of the clinician* as an indispensable component of EBM (Sackett, 1997):

> "Good doctors use both individual clinical expertise and the best available external evidence, and *neither alone is enough* (emphasis added). Without clinical expertise, practice risks becoming tyrannized by external evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient."

In addition to external evidence and factoring in the judgment of the clinician, one must recognize a *third* element: the attributes of the actual patient. The significance of the patient in EBM has been demonstrated by the introduction and growing acceptance of such patient-based outcome measures as the *Health Related Quality of Life Index* and *cost-effectiveness*. Indeed, it has been argued that "the most compelling and growing" component of EBM is *the empowerment of the patient in the decision-making process* (Fisher and Wood, 2007). With patients being the best judge of values, clinical decisions are becoming recognized as necessarily shared between the patient and clinician (O'Connor, 2001). Thus, within the past 20 years, the most accurate portrayal of EBM constitutes the *three* domains just discussed rather than simply the pyramid of external evidence sometimes—and erroneously—used as the sole yardstick with which to direct treatment or payments for treatment. By the year 2000, EBM was officially recognized as "the integration of best research evidence with clinical expertise and patient values" (Sackett et al., 2000).

These shifting sands of EBM echo what a few years ago appeared to be a revolutionary upheaval suggested by Wayne Jonas, who presented what appeared to be a recasting of the classical evidence pyramid. In Jonas' presentation of the "evidence house", such entities as use testing, public health, and audience preferences gained ascendancy (Jonas, 2001). It is here that the absolute hierarchy of research methods is abandoned, taking into account the aforementioned values of the patient and practitioner with recognition extended for the first time to such areas of research as laboratory findings in the basic sciences, epidemiology, qualitative case reports, and health services research. Greater accessibility to research information is afforded by delineating the *purpose* for which the research is being used, dividing its presumed *applicability* into three categories (effects, validity, and use). Jonas' approach laid the foundation for what emerged over a decade later as *comparative effectiveness research*, comparing one diagnostic or treatment option to one or more others and encompassing the medical, economic, and social implications of the application and emulation of an intervention used to promote health. Now, such varied audiences as consumers, clinicians, purchasers and policy makers alike could make informed decisions to improve healthcare at both the individual and population levels (Institute of Medicine, 2009). Jonas' more inclusive schematic of the meaningful elements of EBM is shown in Figure 1b.

While these revisions may have addressed the intrinsic, structural problems of EBM, they have not been able to overcome some of the most severe *misuses* of EBM. When EBM is applied in a unilateral, heavy-handed manner, it has run the risk of becoming a "regime of truth" in such a manner as to discourage free inquiry. Put in other terms, it is questionable whether certain applications of EBM promote the multiple ways of knowing considered to be important in most health disciplines (Holmes et al., 2006), falling under a spell which Foucault has referred to as a "clinical gaze", defined as "a regimented and institutionalized version of truth" in which pluralistic inquiry has been suppressed (Foucault, 1973). Real-life concerns, such as cost-effectiveness to the third party payer or prevalence to the epidemiologist have been overlooked. For these reasons, proponents of EBM have taken somewhat of a political retreat, falling back to a position in which the best evidence is now considered to *guide* or *inform* rather than mandating a clinical decision (Glasziou, 2005; Rycroft-Malone, 2008).

Deconstructing the elements of the EBM pyramid

Randomized controlled trials

One might assume that the most controlled investigation (the clinical trial) would yield the most useful information. Indeed, the clinical trial has been referred to as the "*gold standard*" in clinical research (Report of the US Preventative Services Task Force, 1989). But paradoxically, because the double-blind study is so controlled, this most rigorous member of the clinical research hierarchy presents its own difficulties in its limited generalizability. It is clear from the following criticisms that both methodological and translational issues have become bound together:

1. The characteristics of its own experimental patient base (including comorbidities) may differ significantly from those of the individual presenting complaints in the doctor's office.
2. Potentially important ancillary treatments are restricted, screening out conceivably significant and perhaps unidentified elements that occur in the natural setting of the patient's visit to the physician.
3. Outcome results chosen may not necessarily be those used to evaluate a patient's welfare under care of an actual physician.
4. Experimental groups may not be large enough to reach statistical significance, even though the clinical effect may be real in many individuals.

Like any construct, systematic reviews and meta-analyses can only draw its strength from its component parts. But the numerous flaws in those component parts (especially randomized controlled trials) as reviewed in this communication present a significant challenge to EBM and have been reviewed elsewhere (Hawk et al., 2007; Rosner, 2003; Verhoef et al., 2005), with particular relevance to physical medicine.

Systematic reviews, meta-analyses

In an effort to filter out low-quality studies, rating systems of trial quality have proliferated in the attempt to assure that the edifice of EBM used to warrant a therapeutic approach is more than a house of cards. Such rating systems are the cornerstones of both systematic literature reviews and meta-analyses.

A multiplicity of scoring systems for trial quality exists, but what appears to be their arbitrariness rather than objectivity is perhaps best reflected by the rating chart shown in Table 2. It is taken from a recent blend of narrative and systematic reviews by numerous clinical chiropractic researchers headed by Bronfort et al. (2008). What is perplexing is that two out of the eight criteria for quality recognize *blinding of the patient or practitioner* as an attribute for quality ratings. As indicated previously, successful blinding of these parties in any trial involving physical interventions is virtually impossible.

Table 2 Scoring criteria for randomized clinical trials (Bronfort et al., 2008).

1. Similarity of baseline characteristics or adjusted effects reported
2. Concealment of treatment allocation
3. Blinding of patients
4. Blinding of provider/attention bias
5. Blinding of assessor/unbiased outcome assessment
6. Dropouts reported and accounted for in analysis
7. Missing data reported and accounted for in analysis
8. Intention-to-treat analysis/balanced cointervention

1.0 point awarded for YES rating.
0.5 point awarded for PARTIAL rating.
0.0 point awarded for NO rating.
Quality score is calculated by dividing point total by 8 and multiplying the result by 100 to create a 100-point scale.

Even though systematic reviews and meta-analyses are intended to paint as much of a unified picture as possible from disparate pieces of scientific literature, conflicting reports and insufficient detail may yet pull them in opposing directions. This phenomenon is illustrated in Table 3 (Bronfort, 2006).

Taking these varied methodologies and outcomes together demonstrates how averaging disparate approaches obscures vital information and creates the impression that the entire realm of spinal manipulation may be ineffective. A more meaningful approach would be to determine the *distribution* of pain scales among the different cohorts of patients, which uncovers patient responses unseen by simply taking mean differences. As put in other terms by Bronfort, one needs to consider *effectiveness* as well as efficacy, keeping in mind how spinal manipulation may benefit specific patient groups. This may account for the differing conclusions shown above.

Bias and omissions in other systematic reviews and meta-analyses

Regarding their clinical relevance, the very basis of meta-analyses has to be closely scrutinized. One report has gone so far as to compare meta-analyses to statistical alchemy due to their intrinsic nature:

> "…the removal and destruction of the scientific requirements that have been so carefully developed and established during the 19th and 20th centuries. In the mixtures formed for most statistical meta-analyses, we lose or eliminate the elemental scientific requirements for reproducibility and precision, for suitable extrapolation, and even sometimes for fair comparison" (Feinstein, 1995).

Specifically, the following are some of the deficiencies of meta-analyses, most having to do with the sloughing of important clinical information (Feinstein, 1995):

1. Disparate groups of patients of varying homogeneity across different studies are thrown into one analysis, often called a "mixed salad";

Table 3 Conflicting systematic reviews (Bronfort, 2006).

a. For back pain	
Assendelft (Assendelft et al., 2004)	Bronfort (Bronfort et al., 2004)
To assess if SMT is better than anything else	To assess if SMT is an effective treatment
Statistical pooling using unusual method	Statistical pooling not possible
SMT only equal to sham therapy or therapy judged to be ineffective or even harmful	SMT offers more short-term relief than mobilization or detuned diathermy
b. For neck pain	
Gross (Gross, 2004)	Bronfort (Bronfort et al., 2004)
No attempts to isolate SMT	Isolated effect of SMT
Statistical pooling	Statistical pooling not possible
Evidence did not favor SMT or mobilization alone or combined with other physical agents	Chronic SMT/mobilization is superior to general practice for short-term pain
c. For headache	
Astin (Astin, 2002)	Bronfort (Bronfort, 2001)
Focus on comparison with placebo control	Focus on relative efficacy
Statistical pooling	Statistical pooling not possible
Data available do not support SMT	Several treatments including SMT may be effective

2. The weighting of studies of different quality may be inaccurate or absent altogether;
3. One needs to know about the real-world effects in the presentation and treatment of patients; in particular (a) the severity of the illness, (b) comorbidities, (c) pertinent co-therapies, and (d) clinically relevant and meaningful outcomes;
4. Inconsistent statistical techniques are present, pertaining to increments, effect size, correlation coefficients, and relative risk and odds ratios.

To make matters worse, a recent report involving four medical areas (cardiovascular disease, infectious disease, pediatrics, and surgery) indicates that individual quality measures were not reliably associated with the strength of the treatment effect in 276 RCTs analyzed in 26 meta-analyses (Balk et al., 2002).

The fact that arbitrariness and bias can not only creep into but actually *dominate* in meta-analyses is both convincingly and dramatically demonstrated in a recent publication in the *Journal of the American Medical Association*. In their efforts to compare two different preparations of heparin for their respective abilities to prevent post-operative thrombosis, Juni and his colleagues have revealed that diametrically opposing results can be obtained in different meta-analyses, depending upon which of 25 scales is used to distinguish between high- and low-quality RCTs. The root of the problem is evident from the variability of weights given to three prominent features of RCTs (randomization, blinding, and withdrawals) by the 25 studies which have compared the two therapeutic agents. In one investigation, for example, a third of the total weighting of the quality of the trial is afforded to both randomization and blinding, whereas in another, *none* of the quality scoring is derived from these two features. Widely skewed intermediate values for the three aspects of RCTs under discussion are apparent from the 23 other scales presented. The astute reader will immediately suspect that sharply conflicting conclusions might be drawn from these different studies — and these fears are amply borne out by the Forest Plot presented in the study. Here each of the meta-analyses listed resolve the studies they have reviewed into high- and low-quality strata, based upon each of their scoring systems. It can be seen that ten of the studies selected show a statistically superior effect of one heparin preparation over the other, but only for the *low*-quality studies. Seven other studies reveal precisely the *opposite* effect, in which the *high*- but not the low-quality studies display a statistically significant superiority of low-molecular weight heparin (Juni et al., 1999). Depending upon which scale one uses, therefore, one can either demonstrate or refute the clinical superiority of one clinical treatment over the other. In this manner, therefore, all the rigor and labor-intensive elements of the RCT and its interpretation by the meta-analysis are simply reduced to the subjective and undoubtedly capricious human element of value judgment through the arbitrary assignment of numbers in the weighting of experimental quality.

Assigning a high priority to blinding in rating the methodological quality of clinical trials turns out to be especially problematic when any aspect of physical medicine is involved, for the simple reason as stated earlier that it is virtually impossible to blind practitioners and patients in such instances. Consequently, the act of penalizing such trials in their quality ratings due to lack of blinding is counterproductive. Yet, this is precisely what has occurred in systematic reviews addressing a variety of interventions in physical medicine. Specifically, the Jadad rating scale which scores blinding as one of the most influential determinants of clinical trial quality (Juni et al., 1999) was incorporated (and thereby would have excluded meaningful studies) in systematic reviews addressing reflexology (Ernst, 2009) and osteopathy (Posadzki and Ernst, 2011). Lack of blinding was also cited as a flaw in clinical trials rated in a systematic review of physiotherapy for patients with soft tissue shoulder disorders (Van der Heijden et al.,

1997). The fact that such a multiplicity of rating scales has existed in meta-analyses (Juni et al., 1999) and that even the more recent 8-point rating scale (Bronfort et al., 2008) continues to promote blinding even in physical medicine as a criterion of quality suggests that a consensus is lacking as to which rating scale should be uniformly applied, and that a degree of subjectivity and personal judgment continues to prevail.

The fact that meta-analyses and systematic reviews can be taken down to subjective values and scales and generate conflicting results poses a substantial threat to the so-called objectivity of systematic reviews, among them those by Ernst and Canter (2006) and yet another recent study by Walker et al. (2011).

The former review has been criticized by musculoskeletal researchers for its including studies whose criteria could be considered to be unsystematic and biased (Breen et al., 2006; Byfield and McCarthy, 2006; Lewis and Carruthers, 2006; Moore, 2006). Walker's investigation, on the other hand, concluded that combined chiropractic management for low back pain only "slightly" improved pain and disability in the short term, and that there was no evidence to support or refute the assertion that combined chiropractic treatment provided a "clinically meaningful" difference in pain and disability in people with low back pain. From a list of 3699 studies initially sampled on the basis of titles and abstracts, the Walker report selected 168 for full-text review and ultimately included just 12 — at least one of them flawed — and excluded others in which chiropractors delivered the intervention in multiple study arms. By so doing, it may have presented a negatively skewed assessment of the combined chiropractic management of low back pain. The authors indicated how their study was designed to hew more closely to actual clinical practice, yet its design precluded such important considerations in clinical practice as (a) comparative side effects and the relative safety of different interventions, (b) omission of significant clinical information in systematic reviews and meta-analyses, and (c) studies such as the UK BEAM trial which yielded substantially more positive outcomes regarding back pain in which chiropractors played a significant role (UK BEAM Trial Team, 2004).

Rebuilding the edifice of EBM with modified components

As alternatives to alleviate the limited information available from randomized clinical trials, systematic reviews, and meta-analyses, modified designs based upon a blending of observational and experimental studies have been proposed (Verhoef et al., 2005). All attempt to varying degrees to circumvent the problematic issue of blinding in physical medicine:

Pragmatic trials, in which the intervention is intended to represent "real-world" care to enhance the external validity.

Factorial designs, comparing single modalities to a combination of modalities to allow for the assessment of multiple interaction effects between treatments as well as the synergistic effects of two or more interventions.

Preference trials, in which participants with no treatment preferences are randomized as usual, but those with distinct preferences in their care receive their favored treatment, allowing for the assessment of interactions the outcome and treatment preference.

n-of-1, a single patient trial with multiple crossovers between a treatment period and a placebo or standard treatment period, addressing the critique that a conventional randomized controlled trial does not provide information about the individual, but only an average effect of treatment upon a group of individuals.

One other important modification of the traditional randomized controlled trial has been proposed (West et al., 2008):

Randomized encouragement design: To encourage adherence to the trial protocol, participants are either given strong incentives that are outside usual practice or are allowed to choose or decline a specific treatment to which they have been assigned. These particular practices are distinguished from the conventional randomized trial only by the degree of incentivization, or encouragement to drop out if the assigned treatment is not to one's liking.

Taking the modifications of randomized controlled trials a step further, various researchers have proposed:

Pragmatic clinical trials (PCTs), which ask practical questions about the risks, benefits, and costs of intervention as they would occur in routine clinical practice. In addition, they include a diverse population of study participants, recruiting from a variety of practice settings and collect data from a broad range of health outcomes. The interventions which they select are clinically relevant (Tunis et al., 2003).

Whole systems research (WSR), which uses observational studies and includes qualitative as well as quantitative research. In so doing, it provides the opportunity to assess the meaning that patients attribute to an intervention, probing the process and context by which healing occurs. This includes the philosophical basis of treatment, patients, practitioners, practice setting, and methods and materials used as well as family, community, and environmental characteristics and perspectives. Outcomes which are relevant to the patients are selected, and the approach explores how the intervention fits with a patient's life (Verhoef et al., 2005). In so doing, it reveals the role that *expectations* may play in healing (Verhoef et al., 2002). Essentially, WSR seeks to describe the effectiveness of the entire clinical encounter rather than simply a single procedure (Hawk et al., 2007).

Patient-oriented evidence that matters (POEM): This was developed specifically for primary care physicians in 1994 by David Slawson, M.D., and Allen Shaugnessy, Pharm D. Its intention is to focus only on what is important, allowing physicians to disregard much of the medical literature in favor of useful information at the point of care that would assist them in caring for their patients. In much the same way that infectious disease practice is considered the practical application of microbiology, explains Shaughnessy, POEMs represent the mastery of EBM information on a broad, practical platform rather than simply the passive application of EBM (White, 2004).

All these proposed modifications simply restate the aforementioned contention that *clinical experience* as

well as scientific evidence must be taken into a more encompassing and meaningful application of evidence-based medicine (Sackett, 1998). Steps in this direction appear to have been achieved in a systematic review of musculoskeletal physical therapy, which extended beyond the usual confines of systematic reviews by incorporating observational studies, surveys, and qualitative studies as well as clinical trials (Hush et al., 2011).

Taking a different track from traditional meta-analyses and systematic reviews in summarizing the research literature, Cuthbert and Goodheart distilled the results of 12 randomized clinical trials, 26 prospective cohort studies, 17 retrospective studies, 26 cross-sectional studies, 10 case control studies and 19 case reports in their analysis of manual muscle testing (MMT), a procedure which lies at the heart of applied kinesiology (AK). They provided a multiplicity of definitions of reliability (intra- and interobserver) and validity (construct, convergent and discriminant, concurrent, predictive, and emerging construct) of MMT. In so doing, they offered an alternative yet by many criteria just as formal an approach as the statistical methodology of meta-analyses in supporting the credibility of AK and clarifying different modes of muscle testing in the process (Cuthbert and Goodheart, 2007).

Summary and conclusions

In summary, concepts of EBM have been undergoing continuous revisions and need to remain flexible in order for the most effective healthcare decisions to be made. It is the successful fusion of evidence from the literature, clinical judgment from the physician, and expectations and values of the patient which ultimately will determine the best course of therapy and which is currently being rediscovered and readmitted into the pantheon of EBM. Because all forms of quantitative experimentation can never be construed as absolute means of knowing by their reliance upon statistical probabilities—which are culminated in meta-analyses—one must keep in mind the importance of allowing *multiple* forms of inquiry and experimental design to effectively answer a research question. This ecumenical approach to EBM should, in the final analysis, yield the most successful outcomes in healthcare.

Numerous flaws in the designs and executions of systematic review, meta-analyses, and randomized controlled trials as integral (and sometimes exclusive) components of EBM demand that what is commonly regarded as definitive evidence for a healthcare intervention be revisited. Specifically, numerous examples of these designs drawn from the literature have been found to be susceptible to personal bias. As such, their findings are open to debate and are thus less conclusive than first assumed. Intrinsic in the performance of RCTs is the fact that large placebo effects can sometimes obscure the consequences of the verum treatments commonly seen in clinical practice.

Accordingly, alternative research designs and rankings of evidence have been proposed to accommodate such important elements as basic research, epidemiological considerations, qualitative case studies, clinical judgment, comorbidities, and patient values that have been overlooked in traditional schemes of EBM and can be incorporated in more progressive models. The elements of clinical judgment and patient values, having been proposed as two critical components of EBM in addition to research results summarized in systematic reviews and meta-analyses, have traditionally been neglected. For manual therapy, in particular, the implications of these questions are sufficient to welcome numerous proposals for rethinking what is meant by EBM so that it is better equipped to merit its application to healthcare alternatives as envisioned by numerous audiences.

Competing interests

AR is the Research Director of the ICAK-USA.

References

Aravanis, C., Luisada, A.A., 1956. Results of treatment of angina pectoris with choline theophyllinate by the double-blind method. Ann. Intern. Med. 44 (6), 1111—1122.

Assendelft, W.J.J., Morton, S.C., Yu, E.I., Suttorp, M.J., Shekelle, P.G., 2004. Spinal manipulation and mobilization for low back pain. Cochrane Database Syst. Rev. 1, CD000447.

Astin, J.A., Ernst, E., 2002. The effectiveness of spinal manipulation for the treatment of headache disorders: a systematic review of randomized clinical trials. Cephalalgia 22, 617—623.

Balk, E.M., Bonis, P.A.L., Moskowitz, H., Schmid, C.H., Ioannidis, J.P.A., Wang, C., Lau, J., 2002. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. J. Am. Med. Assoc. 287 (22), 2973—2982.

Benson, K., Hartz, A.J., 2000. A comparison of observational studies and randomized, controlled trials. N. Engl. J. Med. 342 (25), 1878—1886.

Breen, A., Vogel, S., Pincus, T., Foster, N., Underwood, M., on behalf of the Musculoskeletal Process of Care Collaboration, 2006. Systematic review of spinal manipulation: a balanced review of evidence?. Letters to the Editor. J. R. Soc. Med. 99, 277.

Bronfort, G., 2006. Conflicting Evidence: What's the Problem? American Association of Chiropractic Colleges/Research Agenda Conference XI, Vienna, VA, March 17, 2006.

Bronfort, G., Assendelft, W.J.J., Evans, R., Haas, M., Bouter, L., 2001. Efficacy of spinal manipulation for chronic headache: a systematic review. J. Manipulative Physiol. Ther. 24, 457—466.

Bronfort, G., Haas, M., Evans, R.I., Bouter, L.M., 2004. Efficacy of spinal manipulation and mobilization for low back pain and neck pain: a systematic review and best evidence synthesis. Spine J. 4, 335—356.

Bronfort, G., Haas, M., Evans, R., Kawchuk, G., Dagenais, S., 2008. Evidence-informed management of chronic low back pain with spinal manipulation and mobilization. Spine J. 8, 213—225.

Byfield, D., McCarthy, P.W., 2006. Flaws in the review. Letters to the Editor. J. R. Soc. Med. 99, 277—278.

Concato, J., Shah, N., Horwitz, R.I., 2000. Randomized, controlled trials, observational studies and the hierarchy of research designs. N. Engl. J. Med. 342 (25), 1887—1892.

Coulter, A., Entwistle, V.A., Gilbert, D., 1999. Sharing decisions with patients: is the information good enough? BMJ 318, 318—322.

Cuthbert, S.C., Goodheart, G.J., 2007. On the reliability and validity of manual muscle testing: a literature review. Chiropr. Osteopat. 15, 4.

Ernst, E., 2009. Is reflexology an effective intervention? A systematic review of randomized controlled trials. Med. J. Aust. 191 (5), 263—266.

Ernst, E., Canter, P.H., 2006. A systematic review of systematic reviews of spinal manipulation. J. R. Soc. Med. 99, 189—193.
Feinstein, A.R., 1995. Meta-analysis: statistical alchemy for the 21st century. J. Clin. Epidemiol. 48 (1), 71—79.
Fisher, C.G., Wood, K.B., 2007. Introduction to and techniques of evidence-based medicine. Spine 32 (19S), S66—S72.
Foucault, M., 1973. The Birth of the Clinic: An Archaeology of Medical Perception. Random House, New York, NY.
Glasziou, P., 2005. Evidence based medicine: does it make a difference? Make it evidence informed practice with a little wisdom. BMJ 330 (7482), 92 (discussion 94).
Gross, A.R., Hoving, J.L., Haines, T.A., et al., 2004. Manipulation and mobilisation for mechanical neck disorders. Cochrane Database Syst. Rev. 3, CD004249.
Gupta, M., 2003. A critical appraisal of evidence based medicine: some ethical considerations. J. Eval. Clin. Pract. 9, 111—121.
Hawk, C., Khorsan, R., Lisi, A.J., Ferrance, R.J., Evans, M.W., 2007. Chiropractic care for nonmusculoskeletal conditions: a systematic review with implications for whole systems research. J. Altern. Complement. Med. 13 (5), 491—512.
Heany, R., 2003. Long-latency deficiency disease: insights from calcium and vitamin D. Am. J. Clin. Nutr. 78, 912—919.
Holmes, D., Murray, S.J., Perron, A., Rail, G., 2006. Deconstructing the evidence-based discourse in health sciences: truth, power and fascism. Int. J. Evid. Based Healthc. 4, 180—186.
http://phpartners.org/tutorial/04-ebph/2-keyConcepts/images/4.2.7.1.jpg. Downloaded 02.18.08.
Hush, J.M., Cameron, K., Mackey, M., 2011. Patient satisfaction with musculoskeletal physical therapy care: a systematic review. Phys. Ther. 91 (1), 25—36.
Institute of Medicine, 2009. National Priorities for Comparative Effectiveness Research. Report Brief. Institute of Medicine, Washington, D.C.
Jonas, W., 2001. The evidence house: how to build an inclusive base for complementary medicine. Western J. Med. 175, 79—80.
Juni, P., Witsch, A., Bloch, R., Egger, M., 1999. The hazards of scoring the quality of clinical trials for meta-analysis. J. Am. Med. Assoc. 282 (11), 1054—1060.
Lewis, B.J., Carruthers, G., 2006. A biased report. Letters to the Editor. J. R. Soc. Med. 99, 278.
Moore, A., 2006. Chair of NCOR on behalf of the members of the National Council for Osteopathic Research. J. R. Soc. Med. 99, 278—279.
O'Connor, A., 2001. Using patient decision aids to promote evidence-based decision making. EMB Notebook 6, 100—102.
Posadzki, P., Ernst, E., 2011. Osteopathy for musculoskeletal pain patients: a systematic review of randomized controlled trials. Clin. Rheumatol. 30 (2), 285—291.
Pronk, N.P., Lowry, M., Kottke, T.E., Austin, E., Gallagher, J., Katz, A., 2010. The association between optimal lifestyle adherence and short-term incidence of chronic conditions among employees. Popul. Health Manag. 13 (6), 289—295.
Reilly, D., 2002. Randomized controlled trials for homeopathy: when is useful improvement a waste of time? Double positive paradox of negative treatments. BMJ 325 (7354), 41.
Report of the U.S. Preventive Services Task Force, 1989. Williams & Wilkins, Baltimore, MD.
Rosner, A., 2003. Fables or foibles: inherent problems with RCTs. J. Manipulative Physiol. Ther. 26 (7), 460—467.
Rycroft-Malone, J., 2008. Evidence-informed practice: from individual to context. J. Nurs. Manag. 16 (4), 404—408.
Sackett, D.L., 1997. Evidence-based medicine. Semin. Parinatol. 21, 3—5.
Sackett, D.L., 1998. Evidence-based medicine. Editorial. Spine 23 (10), 1085—1086.
Sackett, D.L., Straus, S., Richardson, W.S., Rosenberg, W.M.C., Haynes, B., 2000. Evidence-based Medicine: How to Practice and Teach EBM. Churchill Livingstone, Edinburgh.
Tonelli, M.R., 1998. The philosophical limits of evidence-based medicine. Acad. Med. 73 (12), 1234—1240.
Tunis, S.R., Stryer, D.B., Clancy, C.M., 2003. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 290 (12), 1624—1632.
UK BEAM Trial Team, 2004. United Kingdom back pain exercise and manipulation [UKBEAM] randomised trial: effectiveness of physical treatments for back pain in primary care. BMJ 329, 1377.
Van der Heijden, G.J., van der Windt, D.A., de Winter, A.F., 1997. Physiotherapy for patients with soft tissue shoulder disorders: a systematic review of randomized clinical trials. BMJ 315 (7099), 25—30.
Verhoef, M.J., Casebeer, A.l., Hilsden, R.J., 2002. Assessing efficacy of complementary medicine: adding qualitative research to the "gold standard". J. Altern. Complement. Med. 8, 275—281.
Verhoef, M.J., Lewith, G., Ritenbaugh, C., Boon, H., Fleishman, S., Leis, A., 2005. Complementary and alternative medicine whole systems research: beyond identification of inadequacies of the RCT. Complement. Ther. Med. 13, 206—212.
Walach, H., 2001. The efficacy paradox in randomized controlled trials of CAM and elsewhere: beware of the placebo trap. J. Altern. Complement. Med. 7 (3), 213—218.
Walker, B.F., French, S.D., Grant, W., Green, S., 2011. A Cochrane review of combined chiropractic interventions for low-back pain. Spine J. 36 (3), 230—242.
Wallach, H., Jonas, W.B., Lewith, G.T., 2002. The role of outcomes research in evaluating complementary and alternative medicine. Altern. Ther. Health Med. 8 (3), 88—95.
West, S.G., Duan, N., Peguegnat, W., Gaist, P., Des Jarais, D.C., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M., Mullen, D.P., 2008. Alternatives to the randomized controlled trial. Am. J. Public Health 98 (8), 1359—1366.
White, B., 2004. Making evidence-based medicine doable in everyday practice. Fam. Pract. Manag. 51—58.