Integrative Cancer Therapies
http://ict.sagepub.com

Evidence Summaries and Synthesis: Necessary but Insufficient Approach for Determining Clinical Practice of Integrated Medicine?

Ian D. Coulter

Integr Cancer Ther 2006; 5; 282
DOI: 10.1177/1534735406295564

Evidence Summaries and Synthesis: Necessary but Insufficient Approach for Determining Clinical Practice of Integrated Medicine?

Ian D. Coulter, PhD

The heart of evidence-based practice is in fact to be found in the use of evidence gained from systematic reviews or more correctly in the synthesis of evidence from systematic reviews. But just as studies vary in the quality of the design so do systematic reviews, and it is therefore necessary for those wishing to make clinical decisions based on this evidence to evaluate the evidence summaries and synthesis themselves. This article examines the criteria available for evaluating the quality of the evidence summary and synthesis. It provides a set of questions for doing this: *who did the review*; w*hat was the objective of the review*; *how was the review done?* Together these questions allow us to determine the trustworthiness of the review. However, that by itself is insufficient for making clinical decisions. The article suggests that this occurs because the very studies that improve the quality of reviews, that is, the randomized controlled trials, deal with efficacy and not effectiveness. Because they tend to be conducted under ideal conditions, they seldom provide the type of information needed to make a decision vis-à-vis an individual patient. The article suggests that observation studies provide much better information in this regard. The challenge here, however, is to develop standards for judging quality observation studies. In conclusion, systematic reviews and syntheses of evidence are a necessary but an insufficient method for making clinical decisions.

Keywords: *systematic reviews; evidence synthesis; clinical practice*

The definition formulated for evidence-based practice (EBP) is that it is "the conscientious, explicit and judicious use of the current best evidence in making decisions about the care of individual patients. The practice of EBP means integrating individual clinical expertise with the best available external clinical evidence from systematic research."[1] But the heart of EBP is to be found in the last phrase, "evidence from systematic research." Without these reviews, EBP as currently contemplated is not possible, and the most powerful of these reviews involve a synthesis of the evidence. That being the case, it can also be stated that the evidence on which EBP rests is therefore determined by the quality of the systematic reviews and their synthesis of evidence.

DOI: 10.1177/1534735406295564

Systematic Reviews and Synthesis of Evidence

Just as clinical trials vary in their quality, so do systematic reviews. The standard hierarchy of evidence for practice, based on study designs, is the following (from the highest to the lowest): evidence provided by at least one appropriately designed randomized control trial (RCT); evidence provided by a controlled trial that is not randomized; evidence provided by a well-designed cohort or case-control study; evidence provided by a multiple time series; descriptive studies, case reports, and opinions of experts or respected authorities.[2]

Systematic reviews for their part may vary from qualitative reviews where there is no attempt at a synthesis of the findings (they are merely reviewed and presented) to meta-analysis in which the results of independent RCTs are pooled for a total effect size. Synthesis becomes necessary because discrepancies occur between individual RCTs and between RCTs and meta-analysis.[3] In fact, the results of a single, even double-blinded trial can be misleading, particularly if the number of subjects is insufficient to power the study (in effect to give it statistical legitimacy). Meta-analysis overcomes that problem by combining studies that are homogeneous so that the subject pool is larger.

Although the dominant focus of EBP therefore has been the RCT, it is not the single random-based trial that is the gold standard; rather, it is the systematic review of RCTs that is the most significant, particularly those that result in a meta-analysis.

IDC is Director of Clinical Research and Integrative Medicine at Samueli Institute, Corona del Mar, CA. He also holds appointments at UCLA, RAND, and SCUHS.

Correspondence: Ian D. Coulter PhD, Director of Clinical Research and Integrative Medicine, Samueli Institute, 2101 East Coast Highway, Suite 300, Corona del Mar, CA 921625. E-mail: lcoulter@siib.org.

It should be noted that forms of study design other than the RCT may be included in a systematic literature review (nonrandom trials, cohort studies, and simple pre-post case series), but the RCT is given the most weight because it is the design that most clearly establishes efficacy.

The question for clinical practice is whether the synthesis provides sufficient evidence to either do a procedure or stop doing it in terms of the risk and benefit to the patient. In this sense, they are an aid to clinical decision making, not a substitute for it. The quality of the clinical decision, to the extent it is "evidenced based," rests very heavily on the quality of the reviews, and it is therefore necessary to consider how this might be determined.

Assessing the Quality of Systematic Reviews

We can suggest a set of queries that can be asked with regard to systematic reviews.

1. Who Did the Review?

Reviews are done by a wide variety of individuals and institutions. These vary considerably in both their expertise and particularly in the resources they have available to conduct the review. The effect of funding on results has been noted in the literature, and strong and consistent evidence shows that industry-sponsored research draws pro-industry conclusions.[4] The Evidenced Based Practice Centers[5] established in the United States and Canada and funded through the Agency for Healthcare Research and Quality are probably the most generously funded of the institutions doing this kind of work. One other well-known group is the Cochrane Collaboration.[6] But the Cochrane Collaboration is heavily dependent on individuals doing the work on a volunteer basis.

One of the key questions here is, Are there sufficient resources allocated for the review to ensure it was comprehensive? For example, did the systematic review include studies in languages or literature other than English? Also, were studies in English but published in non-English-speaking countries included? In the RAND study of Ayuvedic interventions for diabetes mellitus,[7] of the 73 studies reviewed, 28 came from the Indian literature, and of those on which statistical analysis was performed, 12 were from India and only 8 from Western literature. Clearly, to have excluded the Indian literature would have greatly impacted the results.

Other examples for sufficient resources allocation include the number of databases searched, was "gray" literature included, and were reviewers available who could read the languages in which the articles are written or was the literature translated. These restrictions should usually be found in the inclusion and exclusion criteria of systematic reviews.

2. What Was the Objective of the Review?

Reviews can have a variety of objectives. Clearly one of the most important is to establish efficacy. A second is safety. Where the latter is given some prominence, the search must include case studies, as adverse reactions are overwhelmingly reported by this method. The number of adverse events reported in trials may not be an adequate measure. In the RAND study of manipulation of low back pain, not one single adverse event was reported in more than 30 trials. By using case study reports, however, the authors were still able to calculate a rate.[8] But many reviews exclude case studies. The objective of the review therefore is to determine the nature of the search strategy and ultimately the results.

A review may be done to establish the state of the science. This may be the case when the initial review is unable to identify sufficient trials to conduct a systematic review or a synthesis. This was the case for the NIH consensus conference for caries where only 7 randomized controlled trials were identified for this disease.[9] In this case, the original systematic review was supplemented by a series of literature reviews.

3. How Was the Review Done?

This involves a set of subquestions.

What were the databases searched? Frequently this will be a single database such as PubMed. However, in areas such as integrative medicine or complementary and alternative medicine (CAM), this is clearly insufficient. In a study of antioxidants and cancer,[10] the databases included online library databases, the reference lists of all relevant articles, and other sources such as experts and the personal libraries of project staff and their associates. "Gray" literature was included (abstracts, etc), but the authors did not specifically search for unpublished data. The team also reviewed meta-analyses and systematic reviews. Thirteen biomedical databases were searched: Allied and Complementary Medicine, BIOSIS Previews®, CAB HEALTH®, CANCER LIT®, Cochrane Library, Elsevier Biobase, EMBASE, MANTIS™, MEDLINE®, SciSearch® Cited Ref Sci 1974-1989, Social SciSearch® 1972-2002, SciSearch® Cited Ref Sci.

What were the search terms used? Again this can have a major impact on the results. Using the antioxidant and cancer study, these included the following: vitamin E, vitamin C, and their many pharmacological synonyms (using search terms such as *alpha tocopherol*, *d alpha tocopherol*, *rrr alpha tocopherol*, and *all rac alpha tocopherol*).

The full search strategy for vitamin C was as follows: vitamin c: ascorbic acid (exploded) from Medline, Embase OR ascorbic acid from all other databases OR dehydroascorbic acid* OR ascorbate OR vitamin c OR antiscorbutic vitamin* OR cevitamic acid* AND neoplasms (exploded) from Medline OR malignant neoplastic disease (exploded) from Embase OR (cancer OR neoplasm*) in subject heading field from BIOSIS OR cancer* in title or subject heading field from all other databases OR neoplasm* from all other databases (exception: in cancer literature, the terms for cancer were omitted and just the total of the "vitamin c" terms were used) AND (prevention OR preventive OR therapy OR therapeutic OR treatment) in title, subject heading fields AND human. Although this looks very complicated, without this information you cannot judge whether the search was comprehensive or not.

What were the inclusion and exclusion criteria? In the antioxidant and cancer study, languages other than English were included in the search and the review. The search was limited to human studies and to controlled trials. This information will again be significant if, for example, intravenous vitamin C studies for cancer were excluded.

Who did the reviews? Here you need to know if they were content experts, clinical experts, or systematic review experts, and if a statistician was part of the team. But the question can be broader than simply who reviewed the individual articles; it can refer to who advised them on the search strategy, who reviewed the final literature review for completeness, relevance, and so on. You need to be assured that those doing the review have the expertise to carry out the work.

How was the evidence evaluated? A standard practice is to have at least 2 independent reviewers for the articles and a third to resolve conflicts. You also need to know what information was abstracted from the articles (study design, randomization, blinding, withdrawals, concealment of allocation), were the studies scored using a quality score, what outcomes/endpoints were considered, were secondary outcomes included, and what statistical measures were used.

What synthesis was possible? Where the studies are not homogeneous, it is not possible to use the techniques of meta-analysis. If a meta-analysis was conducted, you need to be assured the criteria of homogeneity was used; you need to know how this was done, and in terms of what variables. Few studies are completely homogeneous, so this always involves some form of judgment.

How was safety evaluated? In using adverse events, assumptions have to be made about what the rate of reported incidents is likely to be. Is there a 10-fold underreporting of adverse events? Is it 20-fold? As noted earlier, if case studies are excluded from the search, adverse advents may not show up very much in the literature. In the case cited earlier on chiropractic manipulation for low back pain, none were reported in the trials but 111 case reports were found in the literature.

Can We Get to Making Clinical Decisions Based on Systematic Reviews?

Yes, no, and it depends. In a recent article, Holmes, Murray, Perron, and Rail (2006) "assert that the evidence-based movement in health sciences constitutes a good example of microfascism at play in the contemporary scientific arena."[11] This statement may seem a little extreme. Examples of misleading meta-analysis have already been documented in the literature.[12] Such a backlash against EBP might reflect that its followers have perhaps claimed too much. However, it does ignore that systematic reviews and EBP have made a significant contribution. There is a surprising lack of evidence that EBP has better outcomes for patients and that medicine or most of it is in fact evidenced based. Some commentators[13] have given figures as low as 15% for medical practices based on any evidence.

But if we consider the alternative to EBP, it is generally thought as contrasted to traditional practice where "emphasis is placed on accumulated knowledge and experience, adherence to accepted standards and the opinion of experts and peers. It is practical, prudent, personal."[14] The problem here is that the opinion of experts has been shown to be the least reliable (and actually has the lowest ranking of all forms of evidence), and ignoring evidence is almost a recipe for therapeutic anarchy (every therapist does his or her own thing) or for the continuation of what is probably the most popular response: doing what one was taught in professional school. There is an ethical issue here in that patients should be able to expect not only the best care but also care that is current with what we know scientifically and that clearly involved knowing the literature.

EBP and systematic reviews have forced the health professions to take note of the literature and to either challenge it or accept it. Systematic reviews do have transparency. The methods used are documented and can be replicated and challenged. They also have the advantage of "clearing the undergrowth" away to highlight where the evidence is problematic, where there are gaps in the evidence, and where practice is clearly at odds with good evidence. They are extremely good at isolating the state of the science; in some cases, one

might say the deplorable state of the science. To this extent, they can help stop the perpetuation of myths instead of medicine.

Two examples from the author's own work might make this clear. The very first meta-analysis on manipulation for low back pain debunked the myth that manipulation had no evidence and no efficacy. In fact, there was a considerable body of evidence in trials, and this showed that for acute low back pain, this evidence was positive.[15] Furthermore, 2 expert panels were held to derive a set of indications for the appropriateness of manipulation for the treatment of low back pain. One panel was multidisciplinary and the other an all-chiropractic panel. Both received the literature review. In the multidisciplinary panel, at first there were some medical members who thought there was never an occasion when manipulation was appropriate whereas there were manipulators who thought any occasion was appropriate. It became apparent that when evidence was presented, both groups moved.[16] Those opposed moved toward manipulating where the evidence was strong, and those in favor of manipulation moved to restrict the manipulation where the evidence was contrary and/or weak. The experiment demonstrated one of the great strengths of systematic reviews: they cannot be ignored. Systematic reviews force a debate about both the evidence and the practice. Those opposed to the findings are forced to challenge them. This allows clinical acumen and experience to become part of the debate. Adhering to a procedure after such a debate is significantly better than blind adherence based on tradition or the opinion of experts.

The Problem of the Randomized Controlled Trial and Practice

From the point of view of clinical practice, one of the major weaknesses of systematic reviews is its heavy reliance on RCTs. Unfortunately, such studies generally test a therapy under ideal conditions and often with homogeneous populations to ensure comparability of the groups when comparing outcomes. But EBP ultimately requires therapies that can be applied in normal practice, that is, effectiveness studies.[17] Although on logical grounds a therapy without any efficacy will not be effective, a therapy that has efficacy may not have effectiveness when applied to heterogeneous populations and normal practice conditions. Furthermore, therapies with equal, or comparable, efficacy may differ considerably in terms of effectiveness.

In contrast, however, RCTs test therapies under ideal conditions[18] and therefore do not often help with determining effectiveness in everyday practice, as opposed to efficacy in a controlled, and usually perfect, setting. There are some very strict ethical limitations to conducting clinical trials that prevent certain populations from participating. If there is a very high risk but low benefit for a subgroup of patients, this might militate against them being included (such as patients with high comorbidities). Conversely, some low-risk patients may not be included because too large a number would be needed to be enrolled to make the study feasible.[19] The end result therefore is that clinically it is not possible to know if the therapy can be applied to groups that were not included in the trial.

Although providers do treat populations, they treat them one at a time. RCTs seldom contain the "soft data" about individual variations, particularly in response to therapy. The type of clinical detail essential for a provider to decide if a given patient is a candidate for a drug, procedure, or therapy is seldom provided in an RCT.[20] They provide the results of average patients, and even then it is an average of those who meet the inclusion criteria. This problem can be solved through observation studies.

Observation Studies and Practice

There is a dilemma about the role of observational studies. On the one hand, they may seem more clinically relevant and include the populations and subpopulations of interest to the health provider, but on the other hand, they do not provide the type of definitive evidence that might persuade the provider to recommend the procedure to the patient. Despite this ambivalence, observation studies continue to be widely published.[21]

There is an expanding body of literature on studies examining RCTs and observational studies for the same disease and intervention.[22] Earlier studies concluded that nonrandomized studies overestimated treatment benefits. More recently, the studies show that both randomized and nonrandomized studies yield very similar results. Some studies[23] have found little evidence of a difference in the treatment effects between RCTs and observational studies when the comparisons were made for the same treatment.[24] The results of well-designed observation studies did not systematically exaggerate the magnitude of the effects of treatment compared to the RCTs for the same topic.

Because of the inclusion and exclusion criterion used in RCTs, an observational study is more likely to include a broader representation of the population. If the comparison is made with cohort and case control studies, the superiority of RCTs is not so clear. Stroup et al[25] put forward a proposal for improving the reporting of meta-analyses of observational studies in an attempt to solve this question of the quality of observation studies. "The popular belief that only randomized, controlled trials produce trustworthy results

and that all observational studies are misleading does a disservice to patient care, clinical investigations, and the education of health care professionals."[24]

Conclusion

In conclusion, systematic reviews and syntheses of evidence are a necessary but an insufficient method for making clinical decisions. Like all methods, they have limitations and can be critiqued on numerous grounds. On the other hand, they represent an advance in reviewing evidence and as they evolve will continue to aid practice. As the work continues around the issues of evaluating other forms of evidence, particularly for observation studies, their relevance will become even more significant. Also as we move from efficacy studies to effectiveness, their contribution might become more relevant to actual practice. That those who developed and support systematic reviews have often made exaggerated claims for the method is clearly true. That they can be abused is equally true. But those who oppose them may also be guilty of exaggeration and run the risk of throwing out the baby with the bathwater.

Acknowledgment

I wish to acknowledge the assistance of Raheleh Khorsan for her help in the preparation of this article.

References

1. Sackett DL. Evidence-based medicine [Editorial]. *Spine*. 1998; 23(10): 1085-1086.
2. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268:2420-2425.
3. LeLorier J, Gregoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337(8):536-542.
4. Krimsky, S. The funding effect in science and its implications for the judiciary. *J Law Policy*. 2005;13(1):43, 59.
5. Participating AHRQ Evidence-based Practice Centers. February 2006. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://www.ahrq.gov/clinic/epc/epcenters.htm. Accessed September 18, 2006.
6. The Cochrane Collaboration. *Evidence Health Care*. Available at: http://www.cochrane.org/docs/ebm.htm. Accessed September 18, 2006.
7. Hardy M, Coulter I, Venuturupalli S, et al. Southern California Evidence-Based practice Center/RAND. *Ayurvedic Interventions for Diabetes Mellitus: A Systematic Review*. Evidence Report/Technology Assessment Number 41. AHRQ Publication No. 01-E040. Rockville, MD: Agency for Healthcare Research and Quality/DHHS. September 2001.
8. Coulter ID. Efficacy and risks of chiropractic manipulation: what does the evidence suggest? *Integrat Med*. 1998;1(2):61-66.
9. Coulter ID. The NIH Consensus Conference on diagnosis, treatment and management of dental caries throughout life: process and outcome. *J Evidence-Based Dental Pract*. 2001;1(1): 58-63.
10. Coulter ID, Hardy ML, Morton SC, et al. Antioxidants vitamin C and vitamin e for the prevention and treatment of cancer. *J Gen Intern Med*. 2006;21(7):735-744.
11. Holmes D, Murray SJ, Perron A, Rail G. Deconstructing the evidence for evidence-based discourse in health sciences: truth, power and fascism. *Int J Evidence-Based Healthcare*. 2006;4: 180-186.
12. Egger M, Smith GD. Misleading meta-analysis [Editorial]. *Br Med J*. 1995;310:752-754.
13. Smith R. Where is the wisdom? The poverty of medical evidence. *BMJ*. 1991;303:798-799.
14. Niedrman R, Badovinac R. Tradition-based dental care and evidenced-based dental care [Editorial]. *J Dent Res*. 1999;78(7): 1288-1291.
15. Shekelle P, Adams A, Chassin M, Hurwitz E, Phillips R, Brook R. The appropriateness of spinal manipulation for low-back pain: project overview and literature review. RAND Report R-4025/1-CCR/FCER; 1991.
16. Coulter ID, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi- and single-disciplinary panel. *Health Services Res*. 1995;30:577-591.
17. Coulter ID. Evidence-based dentistry and health services research: is one possible without the other? *J Dent Educ*. 2001;65(8):714-724.
18. Glasziou P, Guyatt GH, Dans AL, Dans LF, Straus S, Sackett DL. Applying the results of trials and systematic reviews to individual patients [Editorial]. *ACP J Club*. 1998;129(3):A15-A16.
19. Radford MJ, Foody JM. How do observational studies expand the evidence base for therapy? *JAMA*. 2001;286(10):1228-1230.
20. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine" *Am J Med*. 1997;103:529-535.
21. Ray JG. Evidence in upheaval. Incorporating observational data into clinical practice. *Arch Intern Med*. 2002;162:249-254.
22. Ioannidis JPA, Haidich A, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286(7):821-830.
23. Benson K, Hartz A. A comparison of observational and randomized, controlled trials. *N Engl J Med*. 2000;342(25):1878-1886.
24. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25):1887-1892.
25. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology. A proposal for reporting. *JAMA*. 2000;283(15):2008-2012.