

When the best is the enemy of the good –  
The nature of research evidence used in systematic reviews and guidelines

Marcel P.J.M. Dijkers Ph.D., FACRM

For the Task Force on Systematic Review and Guidelines Sponsored by the  
National Center for the Dissemination of Disability Research (NCDDR)

Members of the Task Force on Systematic Review and Guidelines at the time this analysis was prepared included Michael Boninger M.D., Tamara Bushnik Ph.D., Peter Esselman M.D., Steven Gard Ph.D., Wayne Gordon Ph.D., Allen W. Heinemann Ph.D., Mark Sherer Ph.D., David Vandergoot Ph.D., and Michael L. Wehmeyer Ph.D., all of whom contributed to the development and critical review of the ideas presented in this communication. Careful review and suggestions by Mark Johnston Ph.D. and Jean Ann Summers Ph.D. contributed to this statement's final format. Joann Starks of the National Center for the Dissemination of Disability Research provided administrative support to the Task Force.

UNDER REVIEW BY A JOURNAL.  
NOT FOR QUOTATION OR FURTHER DISSEMINATION

Abstract

Evidence-based practice, according to authoritative statements by the founders of this approach to health care, involves using the “*best available*” evidence, in addition to clinical expertise, and patient preferences, to make decisions on the care of patients. However, many systematic reviewers read this as “best possible,” and exclude from their reviews any evidence produced by research of a grade less than the highest possible (e.g., the randomized clinical trial [RCT] for interventions), even if that means making no recommendations at all. Voltaire’s comment that “*the best is the enemy of the good*” is applicable here. Rehabilitation would be disadvantaged especially, as it can boast few RCTs, because of its nature. The myopic focus on the “strongest” designs also may steer researchers away from asking: “What is the best design to answer *this* research question?” Lastly, rehabilitation and other clinicians need to know not just *which* interventions are effective, but also *how* these need to be delivered; information relevant to this latter aspect of knowledge translation is typically produced using “weak” research designs.

Key words: Rehabilitation; Review Literature as Topic; Review [Publication Type]; Research; Practice Guidelines as Topic

Evidence-based practice (EBP) is an approach to health care professional practice that stresses “the conscientious, explicit, and judicious use of *current best evidence* in making decisions about the care of individual patients. The practice of evidence-based medicine ... means integrating individual clinical expertise with the *best available* external clinical evidence from systematic research.”<sup>1</sup> (Emphasis added.) This is the “authoritative” definition, emulated in such later descriptions as “Evidence-based practice [is] an approach to patient care that incorporates the use of *best evidence* from well-designed studies, a clinician’s expertise, and patient values and preferences.”<sup>2</sup> (Emphasis added.) Other definitions refer to the research literature without the reference to “best.” For instance: “The practice of medicine with treatment recommendations that have their origin in objective tests of efficacy published in the scientific literature rather than anecdotal observations,”<sup>3</sup> or “Practice supported by research findings and/or demonstrated as being effective through a critical examination of current and past practices.”<sup>4</sup>

Since its birth in 1991, evidence-based practice has swept first medicine, next other health care disciplines such as nursing and physical therapy, then other professional fields from education to criminal justice, worrying some that its popularity, at least in lip service, may beget its undoing.<sup>5</sup> However, most rehabilitation practitioners and researchers, whether originally trained prior the emergence of EBP or during the period that it became a force, are more interested in trying to learn the basics of EBP than in considerations of its potential negative impact. While there are variations in descriptions of EBP, most adherents would agree that it involves the following main steps:

1. Pose a clinical question
2. Develop a strategy to find evidence relevant to the question
3. Appraise the evidence, in terms of its relevance to the clinical question, and in terms of the strength of the research that produced it
4. Synthesize the evidence
5. Apply the evidence to practice, taking into account local circumstances and patient values.

Two approaches have developed within this framework. One might be called “bedside EBP,” where a single practitioner, faced with a clinical problem, does a quick search on Medline or another database, based on the abstracts identified rapidly selects what look to be the most relevant and strongest studies, retrieves copies of these papers, and then synthesizes their findings and recommendations and integrates it with clinical expertise and the patient’s circumstances, values and preferences to answer his/her starting question. The process is quick, informal, and usually far from systematic. Most practitioners might take a shortcut to the end result by first talking with a trusted colleague, who may have a broad clinical experience or extensive knowledge of the literature. That may not be the *best* evidence available, but it is fast, presumably targeted, and inexpensive.

A second approach to EBP is that taken by groups of clinicians and researchers who join together to develop, in a particular area of health care, materials that are of benefit to clinicians and others who lack the time (and potentially skills) to take steps 1 through 5 themselves in anything but a cursory manner. These teams evaluate individual papers for publication of EBP-focused digests in one of the many EBP journals that have sprung up (*American College of Physicians Journal Club*, *Evidence-Based Nursing*, *Evidence-Based Communication Assessment and Intervention*, etc.), create critically-assessed topics (CATs), perform systematic reviews, or even use systematic reviews to develop guidelines for practice.

Systematic reviews are systematic in that the evidence is searched for, evaluated, and synthesized in clearly defined steps following a protocol that has been written beforehand. Sometimes protocols are based on specific guidelines such as those of the Cochrane groups<sup>6</sup> or the American Academy of Neurology (AAN).<sup>7</sup> All systematic reviews use a hierarchy of research designs, so as to sort stronger evidence from weaker, based on a positivist view of “evidence.” Sackett in 1989 created the first, simple hierarchy:

1. large randomized trials with clear-cut results
2. small randomized trials with uncertain results
3. non-randomized trials with concurrent or contemporaneous controls
4. nonrandomized trials with historical controls
5. case series with no controls.<sup>8</sup>

This classification, with its ambiguous “large” vs. “small” standard and other problems, now is a historical curiosity. Better hierarchies with from four to up to ten levels have been published, for reviews addressing various types of clinical questions: therapy, screening and diagnosis, prognosis, costs. (Some claim that even the best hierarchies published disregard developments in research methodology over the last 20-40 years. The National Center for the Dissemination of Disability Research (NCDDR) Panel on Standards and Methods of Evidence is expected to publish shortly its recommendations for evidence grading, specifically grading of evidence in disability/rehabilitation research.) The better hierarchies, those of AAN for example, take quality of the research *implementation* as well as basic research *design* into account in differentiating stronger from weaker designs.

In drawing conclusions and making recommendations, the authors of systematic reviews take into account the quality, quantity and consistency of the evidence from many papers and other sources to phrase recommendations. Again, there has been increasing sophistication over time in how this is done. Sackett distinguished three categories of recommendations, differentiated on the basis of a simple “nose count”:

- I. supported by one/more level 1 studies;
- II. supported by one or more level 2 studies;
- III. supported only by level 3, 4 or 5 studies.

Now both quality and consistency, as well as the number of studies and their basic design, may be used to qualify recommendations on a scale ranging from “should/should not be done” through “should/should not be considered” to “may/may not be considered” to “no recommendations.”

Unfortunately, many systematic reviews and guidelines published in recent years have adopted an all-or-nothing approach to the evidence base. Cochrane group reviews may be the most extreme; in many instances only evidence for therapeutic interventions resulting from randomized clinical trials (RCTs) is accepted. If that level of evidence is lacking, “more research” is recommended, and no recommendations for practice are made. Other groups follow a similar practice, although they may draw the line at a different level in the evidence hierarchy. For instance, AAN guidelines specify that no recommendation should be made if there is not at least one Class II study or two consistent Class III studies, and that the recommendation to be made when this minimum level of evidence is available is to be phrased in terms of “may be considered” (or “may not be considered”, as appropriate.)<sup>7</sup>

When a well-respected statistician-methodologist like Douglas Altman goes on record stating “Only randomised trials allow valid inferences of cause and effect. Only randomised trials have the potential directly to affect patient care--occasionally as single trials but more often as

the body of evidence from several trials, whether or not combined formally by meta-analysis”<sup>9</sup> it is not surprising that the misunderstanding spreads in EBP circles that only RCTs can contribute information that is of use in clinical decision making, as reflected in the following: “Treatment decisions in clinical cardiology are directed by results from randomized clinical trials (RCTs).”<sup>10</sup>

It would seem that Voltaire’s comment that “the best is the enemy of the good” (*le mieux est ennemi du bien*) is applicable here. Some systematic review panels or their parent guideline development organizations have raised the bar on the level of evidence required so high that in their reviews no appropriate evidence is discerned, resulting in no recommendation. However, that would appear to go against the grain of EBP as defined by some of its pioneers – as expressed in the quote from Sackett et al. above: “judicious use of *current best evidence* in making decisions.”<sup>1</sup> Similar sentiments can be found in other key EBP texts, such as the book by Straus et al.: “By *best research evidence* we mean valid and clinically relevant research, often from the basic sciences of medicine, but especially from patient-centered clinical research into the ... efficacy and safety of therapeutic, rehabilitative and preventive regimens.”<sup>11</sup> p 1 “Best” should be understood in the meaning of “*best available*,” not as “*best possible*.” By refusing the benefit from whatever value there may be in “flawed” research, the EBP practitioners who refuse to consider anything below a certain evidence grade throw away research that may be informative for the clinical issue in question. Depending on the level of scrutiny applied, they may accept a poorly executed randomized trial over an exemplary case-control study. Secondly, although it would be too bold to state that a panel of reviewers carefully considering meager evidence is *always* more knowledgeable than the lone clinician who has only his or her own experience and possible uncritical reading of the literature on which to rely, in most instances “expert consensus supplemented by weak evidence from the research literature” likely is preferable over “the lone practitioner’s intuition.”

The disregard of “weaker” studies is especially damaging in rehabilitation, because there are so few clinical trials on which to rely.<sup>12</sup> This shortage is due in large part to the nature of rehabilitation: a coordinated treatment effort of many disciplines all using treatments and approaches individualized to the patient, and focusing on long-term outcomes that are affected by multiple personal and environmental factors that largely are not under control of the rehabilitation team. In addition, realistic placebos are not available for many interventions, and blinding (of providers, and sometimes even of patients) is not feasible. (See Johnston et al. for additional issues justifying rehabilitation research’s “low” evidence levels.)<sup>13</sup> Rehabilitation research is not unique in this respect; behavioral medicine, health services research and others share the problem that their treatments do not fit the mold of what often is the exemplar in EBP: the drug vs. placebo short-term double-blinded RCT.

Another consideration is that it is not simply a matter of settling for second-best. The real question is not “What is the most rigorous research design?” but “At this time, what is the best research design *for the research question or practical problem at issue?*” These are not the same. Large RCTs can be premature and can take funds away from the needed development of new interventions. Traditional RCTs apply narrow selection criteria, and therefore their results do not generalize well to a wider universe of patients; “practical clinical trials” have been proposed as a way of producing evidence with more applicability to real life.<sup>14</sup> RCTs are largely inapplicable to assistive technology and environmental modifications, which are core interventions in disability and rehabilitation. In some instances, RCTs are unnecessary, because strong evidence can be generated by means of a much weaker design. For instance, who would do an RCT to test whether wheelchairs work? Clearly, standards for “best research design” in disability and

rehabilitation as in other health care and human services fields cannot be driven by an insistence on large RCTs or an uncritical application of standards from certain evidence-based medicine adherents.

A further issue related to the practice of restricting EBP reviews to RCTs is the wide variations in the interventions that may occur in research in areas such as in rehabilitation, social services and education. In medical research this may not be a problem when the intervention involves a single active ingredient expressed in an easily measured dosage, such as a drug. But in other professional fields, the “intervention” may consist of much more difficult-to-measure entities such as parent training, job coaching or self-advocacy training. When the process of synthesizing the body of evidence about these types of interventions is restricted to RCTs, much useful information that could guide practitioners may be lost. Reaching a judgment about effectiveness of such interventions based on the overall body of evidence often requires selection of studies in which the intervention may have been implemented in many different ways or at many different intensities. The “average effect size across many studies” on which the typical EBP systematic review judgment is based does not provide much guidance for practitioners about how, specifically, to apply the intervention to their own clients or students. In contrast, coupling meta-analysis of RCT studies about a particular intervention with other information gathered from, for example, meta-syntheses of qualitative studies<sup>15</sup> could provide a rich source of guidance for practitioners. If the end goal is the incorporation of best available research into decision making about practices, then for knowledge translation purposes the best that different research approaches have to offer should be included in the synthesis.

No one is likely to claim that RCTs are equal to other designs, at least for demonstrating internal validity – the conclusive proof that a certain intervention has specific positive and negative consequences, compared to placebo or compared to another treatment. (The relative weakness of RCTs versus other designs when it comes to external validity - the generalizing of study findings to a group of which the study subjects are representative - has been discussed extensively in recent literature.)<sup>14, 16</sup> RCTs *are* better, and if designed and executed well they offer a higher level of confidence that a particular treatment is better than or is not significantly different from another treatment or placebo. This level of confidence in a conclusion based on study data cannot be matched by other, observational, designs, however large the sample or however sophisticated measurement of outcomes. However, this gold standard is feasible only in limited circumstances. There are so many treatments and approaches in rehabilitation that deserve evaluation that application of RCTs to them all could exhaust the NIH budget, let alone that of NIDRR. If we are to gain information on treatments that work in rehabilitation (for specific categories of patients, at a particular stage of their disablement career) we need to make creative use of research designs that are less restricted and less expensive than clinical trials.<sup>17</sup>

The argument is not that in all circumstances any level of evidence is better than nothing. If only one small study has been done, of questionable quality, and its findings contradict common sense, there obviously is no reason to base recommendations on those findings. And if a number of very similar studies has been done, some of high quality and some of lesser strength, it is defensible to disregard the latter and base recommendations on the former only – although it should be noted that the filtering of studies based on quality is still a contentious issue.<sup>18-24</sup> While it may take the EBP community some more time to determine under what circumstances quality filtering is or is not recommended, the issue addressed here is one tangentially related: what to do in situations where there is no “embarrassment of riches” when it comes to evidence, but only a few studies are available, all of a design weaker than an RCT or equivalent.

If we are to offer guidance to clinicians as to what approaches may or likely will be most effective or efficient with their patients and clients, our systematic reviews need to be more catholic than allowed by the EBP purists, and sometimes accept, by necessity, all levels of evidence. It is never the case that in the absence of recommendations from a systematic review no rehabilitation services are delivered; given the need to help patients with their impairments and problems, rehabilitation clinicians almost always will try something. If that “something” is supported by weak evidence carefully considered by expert clinicians and researchers, it likely will be better than what a single clinician not guided by the literature will create.

As long as the strength of the evidence is carefully set forth and taken into account along with the quantity and consistency of the evidence, little harm is possible, and much benefit may result. Let’s not make the best the enemy of the good.

## References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. *BMJ*. 1996;312:71-72.
2. Arizona State University College of Liberal Arts and Sciences. ASU College of Nursing leads effort to improve patient care. Available at: <http://www.newswise.com/articles/view/515830/>. Accessed 09/21, 2007.
3. Gerontology Research Group. Glossary for terms in gerontology. Available at: <http://www.grg.org/resources/glossary.html>. Accessed 09/21, 2007.
4. Canadian orthopaedic nurses association. Standards for Canadian orthopaedic nursing. Available at: <http://www.cona-nurse.org/STANDARDS/standards.pdf>. Accessed 09/21, 2007.
5. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine." *Am J Med*. 1997;103:529-535.
6. Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [Undated September 2006]*. Chichester, UK: John Wiley & Sons, Ltd.; 2006; No. The Cochrane Library, Issue 4.
7. Edlund W, Gronseth G, So Y, Franklin G. Clinical practice guidelines process manual - 2004 edition. St. Paul MN: American Academy of Neurology; 2004.
8. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1989;95:2S-4S.
9. Altman DG. Better reporting of randomised controlled trials: The CONSORT statement. *BMJ*. 1996;313:570-571.
10. Hernandez AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: Are most of them misleading? *Am Heart J*. 2006;151:257-264.
11. Straus SE, Richardson WS, Glasziou P, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. Third ed. Edinburgh; New York: Elsevier/Churchill Livingstone; 2005.
12. Johnston MV. Desiderata for clinical trials in medical rehabilitation. *Am J Phys Med Rehabil*. 2003;82:S3-7.
13. Johnston MV, Sherer M, Whyte J. Applying evidence standards to rehabilitation research. *Am J Phys Med Rehabil*. 2006;85:292-309.
14. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290:1624-1632.
15. Sandelowski M, Docherty S, Emden C. Qualitative metasynthesis: Issues and techniques. *Res Nurs Health*. 1997;20:365-371.
16. Horn SD, DeJong G, Ryser DK, Veazie PJ, Teraoka J. Another look at observational studies in rehabilitation research: Going beyond the holy grail of the randomized controlled trial. *Arch Phys Med Rehabil*. 2005;86:8-15.
17. Horn SD, Gassaway J. Practice-based evidence study design for comparative effectiveness research. *Med Care*. 2007;45:S50-7.
18. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. *J Clin Epidemiol*. 2006;59:1249-1256.
19. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol*. 2005;5:19.

20. Cipriani A, Malvini L, Furukawa TA, Barbui C. Relationship between quality of reports of antidepressant randomized controlled trials and treatment estimates: Systematic review, meta-analysis, and meta-regression analysis. *J Clin Psychopharmacol*. 2007;27:352-356.
21. van der Velde G, van Tulder M, Cote P, et al. The sensitivity of review results to methods used to appraise and incorporate trial quality into data synthesis. *Spine*. 2007;32:796-806.
22. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: Implications for the conduct of meta-analyses. *Health Technol Assess*. 1999;3:i-iv, 1-98.
23. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics*. 2001;2:463-471.
24. Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol*. 2001;54:651-654.