ALTERNATIVE THERAPIES

IN HEALTH AND MEDICINE

A PEER-REVIEWED JOURNAL · MAR/APR 2011 · VOL. 17, NO. 2 · **$14.95**

ANTHROPOSOPHIC MEDICINE IN PEDIATRIC PRIMARY CARE · TOPICAL TEA TREE OIL AS WOUND THERAPY · HOMEOPATHIC INJECTION THERAPY FOR MORTON'S NEUROMA · OPERATIONAL DEFINITION OF CAM · COCHRANE CAM REVIEW: MUSIC THERAPY FOR BRAIN INJURY · CONVERSATIONS/BRADLY JACOBS, MD, MPH

03>

0 74470 92705 1

ALTERNATIVE THERAPIES
IN HEALTH AND MEDICINE
WWW.ALTERNATIVE-THERAPIES.COM
MAR/APR 2011, VOL. 17, NO. 2

TABLE OF CONTENTS

REVIEW ARTICLE

CONVERSATIONS

DEPARTMENTS

ABOUT THE COVER

Dressed in vibrant shades of red, a woman spins as she dances with sheer enjoyment. "This composition is all about self-expression, and that is what makes her so beautiful," says an inspired Joseph Adibleku.

Pretty Lady. Acrylic on canvas, 29.9" W x 26.0" H, Joseph Adibleku. Joseph's paintings can be found at novica.com.

ALTERNATIVE THERAPIES IN HEALTH AND MEDICINE (ISSN 1078-6791) is published bimonthly (January, March, May, July, September, November) by InnoVision Professional Media, 1408 Northland Drive, Suite 306, Mendota Heights, MN, Tel: (877) 904-7951, Fax: (651) 344-0774. E-mail: ATHM@innovisionhm.com. *Alternative Therapies in Health and Medicine* is indexed in *Index Medicus, CINAHL, Science Citation Index-Expanded (SciSearch®), ISI (Institute for Scientific Information) Alerting Services, Current Contents®/Clinical Medicine, EMBASE (Excerpta Medica),* and MEDLINE.

For subscription questions please call toll-free: US only, (877) 904-7951; outside the US, (651) 251-9684. Annual individual subscriptions: US and possessions: $68; foreign: $97 (US). Institutional rates: US: $171; foreign: $237 (US). Single copies: US: $12; all other countries: $18 (US). Periodical postage paid at St Paul MN, and additional mailing offices. Postmaster: Send address changes to ALTERNATIVE THERAPIES, PO Box 11292, St Paul, MN 55111. Allow 4 to 6 weeks for change to take effect. The name and title ALTERNATIVE THERAPIES IN HEALTH AND MEDICINE is protected through a trademark registration in the US Patent Office. Printed in the USA.

INNOVISION PROFESSIONAL MEDIA, INC.
1408 Northland Drive, Suite 306 • Mendota Heights, MN • Tel: (877) 904-7951 • Fax: (651) 344-0774 • Web: www.alternative-therapies.com

President & Group Publisher, DICK BENSON • Vice President & CFO, JOHN BENSON • Circulation Director, NICK COLLATOS • IT Manager, SAM BHATT
Administrative Assistant, KELLY SMALL

Advertising Sales
Publisher, DICK BENSON • (651) 251-9617 • dbenson@innovisionhm.com

ALTERNATIVE THERAPIES
IN HEALTH AND MEDICINE

WWW.ALTERNATIVE-THERAPIES.COM

Managing Editor, SUZANNE SNYDER • Creative Director, LEE DIXSON • Associate Editor, ANNE LANCTÔT
E-mail: ATHM@innovisionhm.com • Web: www.alternative-therapies.com

Comparative Effectiveness Research: Does the Emperor Have Clothes?

Ian D. Coulter, PhD

With the recent allocation of a $1.1 billion "down payment" to fund comparative effectiveness research (CER) from the American Recovery and Reinvestment Act of 2009 (generally referred to as the stimulus package) and with $300 million being allocated for the Agency for Healthcare Research and Quality (AHRQ), $400 million for the National Institutes of Health, and $400 million for allocation at the discretion of the Secretary of Health and Human Services and with the National Center for Complementary Alternative Medicine putting out a request for research proposals for Comparative Effectiveness Studies of Complementary and Alternative Medicine, it is safe to say CER has entered a new era. CER solves two historical concerns for complementary and alternative medicine (CAM) researchers; first it focuses on effectiveness not efficacy; second it tests holistic approaches to care. Because it allows the providers to give care in any way they choose, it avoids the problem of reductionism inherent in standard random controlled trials. In CER, the provider can continue to practice holistically and to use individualized medicine to treat the patient. However, amid the largely positive responses to this move among researches in CAM, a more critical evaluation might be in order. This article argues that while the move to effectiveness research is a positive move for CAM, CER as currently being talked about and funded may just be a new form of privileging certain forms of evidence at the expense of other equally important and perhaps more relevant evidence. (*Altern Ther Health Med.* 2011;17(2):8-15.)

Ian D. Coulter, PhD, is a professor at the University of California, Los Angeles; RAND/Samueli Chair for Integrative Medicine and Senior Health Policy Researcher, RAND Corporation, Santa Monica, California; and on the research faculty at Southern California University of Health Sciences, Whittier.

Corresponding author: Ian D. Coulter, PhD
E-mail address: coulter@rand.org

Although comparative effectiveness research (CER) has been around for some time, the recent announcements that the American Recovery and Reinvestment Act of 2009[1] (generally referred to as the stimulus package) would make a "down payment" of $1.1 billion to fund comparative effectiveness research has given a tremendous boost to this form of research.[2] Followed by announcements from the Agency for Healthcare Research and Quality (AHRQ) that it would allocate $300 million[1] and the National Institutes of Health (NIH) will allocate $400 million with $400 million for allocation at the discretion of the Secretary of Health and Human Services, CER is about to enter a new era. It is not that CER is new; AHRQ has been funding and conducting CER. This new funding and focus will "develop a definition, establish prioritization criteria, create a strategic framework, and identify priorities for CER; . . . foster optimal coordination of CER conducted or supported by federal departments; and . . . formulate recommendations for investing the $400 million provided to the Office of the Secretary."[1] In addition to actual research, the funds will be used to train researchers, develop methods, build infrastructure, and disseminate and translate research into practice. We should also keep in mind, as pointed out by Conway and Clancy (2009),[1] the whole field of health services research under which CER falls is only 1.5% of the total for biomedical research health and only 0.1% of the total spent in the United States for health care.

For complementary and alternative medicine (CAM) and integrative medicine (IM) researchers, the fact that the National Center for Complementary Alternative Medicine (NCCAM) has issued a request for research proposals for CER of CAM has added to its importance for the field.[3] Given that research generally follows the funding, we can expect to see a quantum leap in the number of CER trials done in CAM. Adding to this of course is the recommendation about the importance of CER from the prestigious Institute of Medicine (IOM) recommendation for CER in 2009.[4] For the most part, this move to CER has been enthusiastically supported by the CAM community. A recent workshop for Stakeholders Symposium held by the Center for Medical Technology Policy and the Institute for Integrative Health in November 2009[5] showed widespread enthusiasm among those who work in IM and CAM for this type of research.

This possibly reflects the long-held concern in the CAM/IM community about the relevance of randomized controlled trials (RCTs) and the privileging of them for establishing efficacy and the concern about a general lack of effectiveness studies. Efficacy refers to the outcomes attributable to a specific therapeutic intervention tested under controlled and ideal conditions in an RCT. Its strength is that the outcome can be attributed causally to the

intervention. Effectiveness refers to what works in real practice with all the complications that occur in such settings. Here the patients are not preselected and providers practice in the way they would in normal practice. So on one level, the move towards increasing the importance of CER and the increase in funding to conduct CER will be seen as a positive move in the CAM/IM community. But amongst the enthusiasm, this might be a good time to apply a more critical perspective and to make sure we are not simply moving from one form of privileged evidence, the RCT, to another form of privileged evidence, the CER, which has its own set of limitations and problems. That is, to ensure the emperor really does have clothes or at the very least is partially clothed.

WHAT IS COMPARATIVE EFFECTIVENESS RESEARCH?

The IOM definition[4] of CER states that

> *CER is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both individual and population levels.*

AHRQ states that their Effective Health Care Program's purpose is to fund research that provides reliable and practical data that can inform decisions in clinical practice.[1] CER has been identified by several names in the past: pragmatic trials, head-to-head trials, practical clinical trials. They are called pragmatic trials because their purpose is to determine pragmatically what works in practice; hence, they are also referred to as practical trials. A head-to-head trial occurs when two different therapies are compared to one another in a pragmatic trial as opposed to being compared to a placebo or control group as would happen in a RCT. Given the importance being attributed to CER, it is crucial to know exactly what it is.

Some writers have contrasted CER (or practical clinical trials [PCTs]) with explanatory trials.[6] The latter are hypothesis-driven and usually done with the hope of revealing the biological effect of a treatment. In contrast, CER or pragmatic trials are done to assist decision-makers: "PCTs address practical questions about the risks, benefits, costs of an intervention as they would occur in routine clinical practice."[7(p1626)] They further note, "The most distinctive features of PCTs are that they select clinically relevant interventions to compare, include a diverse populations of study participants, recruit participants from a variety of practice settings, and collect data on a broad range of health economics."[7]

Zwarenstein and Treweek [8] note that this is a mismatch between the clinical setting in which decisions must be made and the RCTs (explanatory trials to test hypotheses). They note "evidence from an explanatory trial is unlikely to inform a pragmatic question, nor vice versa." They also note, however, that the bulk of studies have not been pragmatic trials. One review has shown only 100 pragmatic designed randomized trials in the US National Library of Medicine.[8]

Three major reasons are given for this enormous imbalance between the numbers of CER and RCTs. The first is that there is no incentive for private companies who develop either drugs or devices to do comparative studies under less-than-ideal conditions.[8] Usually, the companies want to maximize the chances they get a positive outcome for their products, and secondly, they have no interest in head-to-head trials with competitors. Related to this is that the US Federal Drug Administration (FDA), whose concern is the safety and protection of the public, is more likely to stress animal studies first and Phase I to Phase IV trials but not pragmatic trials.[8] This then has a direct impact on what companies will do. Dominated as it is by scientific interests, NIH for its part is more likely to opt for the more rigorous RCT, explanatory trial and identifying the biological mechanisms involved. The end result is that "Neither of the major sources of funding for clinical research in the United States—the NIH and the medical products industry—has as a primary mission of the goal of ensuring that studies are performed to address clinical questions important to decision makers."[7]

One key feature of CERs is they are focused on effectiveness, not efficacy.[2] Therefore, it is important to examine the difference between these two concepts.[9] Efficacy establishes a causal connection between an intervention and a specific outcome. To create this link, however, it is necessary to control all biases so that the only thing contributing to the outcome is the specific intervention. The most rigorous methodology for doing this is the RCT and in particular, the double-blind RCT that includes a placebo or sham. In this model, it is possible to measure how much of the outcome is placebo effect by studying the arm that got the sham intervention (in effect, a nonintervention). If the patients in this arm get a measurable effect (outcome) but it is less than the intervention arm, it is possible to determine how much of the outcome is due to placebo effect and how much is the result of the therapy. This is the model for most pharmaceutical research and the one required by FDA for approval of most pharmaceutical products and medical devices.[2] Yet achieving accurate results requires that the enrollments in the trial are controlled and random, that the intervention is standardized and controlled (it must be constant and identical across all subjects), that the populations in the trials are homogeneous, and that the outcome measures are standardized and objective. The trials will have at least two arms, one in which the intervention is given and one in which a sham treatment or a placebo is given. Individual participants are randomly allocated to one of the arms. Neither the provider of the therapy nor the patient should know which arm of the trial the patient is in (double blinding).

The problem is that to achieve the kind of controls needed for an RCT, researchers end up with a situation that is not like normal practice where ultimately practitioners want to know if the therapy will work. The exclusion criteria for the trial participants may ensure that the very subpopulations the provider wants to treat were not even included in the trial. Also, it may not be either practical or feasible, economically or otherwise, to

implement the therapy in normal practice. The problem is that the evidence from RCTs may be rigorous but not relevant to the real world of practice. Though observation studies are more relevant, they in turn are less rigorous and may not provide definitive information for making decisions.[10]

Another way of contrasting the two types of trials was given by Schwartz and Lellouch: "[I]n most trials the treatments may be defined in two ways. Either ordinary current practice may be adhered to ('normal' conditions) or more exacting conditions may be introduced which could only be met in the course of a trial ('laboratory' conditions)."[6(p500)]

However, evidence-based practice (EBP) or evidence-based medicine requires outcomes that can be achieved in the average practice with the average provider and with normal patients. Something that has efficacy may not have effectiveness in practice or a low level of effectiveness, and two therapies with equal efficacy may have quite different levels of effectiveness. The therapy that is the most efficacious may in fact have a lower level of effectiveness than an alternative with lower efficacy. Logically it is thought that something with no efficacy will have no effectiveness, but it is certainly the case that something with high effectiveness could have a lower efficacy as determined by RCTs. Because of this, some authors have called for putting the "practice" back in "practice-based evidence."[11] They also have seen CER as a way of doing that.[12]

Formal logic enters into the consideration between the two types of trials in another manner. A negative outcome (effect) in an explanatory trial is thought to be a fatal blow because if there is no effect when the trial is done under ideal conditions, it is highly unlikely that it would hold under less ideal conditions. This is akin to Karl Popper's argument for the logical superiority of falsification over confirmation.[13] A positive effect in an explanatory trial still leaves one with uncertainty about what would happen under less ideal conditions. In pragmatic trials, however, a negative effect leaves one unclear about whether it might work under more ideal conditions and a positive effect can be informative about how it works under normal conditions (effectiveness).[14]

Thorpe et al (2009)[15] prefer to use the term *pragmatic* for an effectiveness trial and *explanatory* for an efficacy trial. The former looks at the effects of an intervention under conditions in which it will be applied in practice. The latter tries to determine the outcome testing under ideal conditions. Those using the terms *explanatory* and *pragmatic* have usually seen the former as

> *highly controlled, with strict eligibility criteria, restrictions in the way the intervention and cointerventions are delivered, and focused on surrogate or biological outcomes. Most methodologists describe pragmatic trials as enrolling all patients to whom health care providers might offer the intervention, allowing clinicians to administer the intervention and cointerventions without restrictions, and measuring patient-important outcomes.[16]*

Tunis, Styer, and Clancy[7] have noted that most systematic reviews of RCTs and observation studies conclude that either there are gaps in the evidence available or that the quality of the evidence is not good. This means that the amount of good evidence and the type of evidence that policymakers, insurance companies, decision-makers, providers, and patients have to make practical decisions is very limited. Again, CERs are seen by some as a corrective to this situation.

In conclusion, Maclure (2009), discussing how to describe pragmatic trials to policymakers, states, "pragmatic trials are real-world studies 'for decision' whereas explanatory trials are specialized studies for 'information.'"[17]

WHAT IS WRONG WITH COMPARATIVE EFFECTIVENESS RESEARCH?

There have been some debate and confusion over the term "pragmatic trials." In some instances, this is not much more than a semantic difference as in the discussion over whether they should be called practical clinical trials[7] or pragmatic clinical trials.[6]

In some instances, however, this is a more serious debate since it is not just a semantic difference but a difference over what is the nature of such trials. This is the case in the debate over mechanistic trials vs practical trials as opposed to explanatory trials vs pragmatic trials. The mechanistic-practical group[18] argues that the explanatory-pragmatic framework confuses the purpose of the trial (its aim) with the structure of the trial and also ignores the varying perspectives of those using the trials to make policy and clinical decisions. Basically, this camp argues that the sort of pragmatic trial used would often be of little use in making clinical decisions:

> *Indeed, a pragmatic trial that enrolled a wide range of patients with varying degrees of depression diagnosed and managed by a mix of primary care practitioners and psychiatrists using a wide variety of cointerventions would be ill suited to inform this particular patient-clinician dyad.[18]*

So a pragmatic trial may not be the best solution for informing real-world decisions. To truly apply results to the real world, the trialists would have to establish what context the results are to be used in prior to performing the trial.

Karanicolas et al[18] think that policymakers' interest is in groups/populations while clinicians' interest is individual patients and that these different interests require different research designs. They feel the current explanatory-pragmatic framework ignores this dilemma. They feel mechanistic trials are ones whose purpose is to address a biological relationship. It is a practical trial when it collects data that bear on health care decisions. How useful it will be in conducting a practical trial will depend on what kind of decision must be made, and the structure of the trial will depend on what kind of decision must be made. Basically, what Karanicolas et al[16] argue is that the term *practical trials* may cover trials with quite different structures (choice of patients, interventions, comparators, permitted coint-

erventions, and outcome measures), whereas the pragmatic trials are usually viewed as enrolling all patients to whom the provider offers care for a particular intervention and allowing the provider to offer both the intervention and cointreventions as they see fit. Karanicolas et al note that this may not be the most pragmatic trial given the decisions that it will be used for.[16]

The full debate between the "mechanists" and the "pragmatists" is beyond our purpose here, but the pragmatists[19] agree that the former have identified an important problem in how explanatory and pragmatic trials have been interpreted. The response of the pragmatists is that the mechanistic-practical continuum moves too many trials from the explanatory end of the continuum over to the practical end and that designing trials to test different specific circumstances is not practical (and can be overcome somewhat by using subgroup analysis), ignores other relationships such as the psychological and sociological, and seems to suggest that clinical decisions should be based on individual trials and not systematic reviews.[19]

Where the two groups agree is that there is a continuum from explanatory to practical/pragmatic and that whether a trial is judged to be more or less practical/pragmatic will depend on the perspective and context: "The perspective of pragmatic trials should be that of (clinical, public health, or health system) policymakers and not that of individual clinicians and patents."[19]

One of the major challenges of pragmatic trials is that to show effects under less-than-ideal conditions, researchers may need a much larger sample size which results in higher costs. Combine this with the fact that at least with manufactured treatments and devices, there is no economic incentive for a company to carry out a CER[8] and no FDA requirement to do so. In industry, 90% of trial funding is for phase I, II, and III trials, which are all required by FDA, and only 10% for phase IV trials.[7]

As noted earlier, NIH has also not been highly motivated to conduct CER, partly because research funded by NIH tends to be explanatory in nature and has no mechanism to identify the priority questions that might be solved by CER.[7] Its funds go overwhelmingly to biomedicine (70%).[7] As Tunis et al (2003) note, even the $300 million earmarked for AHRQ in 2003 would not fund a large expansion in their CER studies if AHRQ was to meet its other core responsibilities.[7]

Tunis et al make several proposals for improving the situation vis-à-vis CERs: establish a system for determining the priorities or gaps in our knowledge; encourage decision makers not only to use high-quality evidence but also to include CER; create operational infrastructures within which to conduct CERs similar to primary-care settings, including training physicians to conduct CER; address methodological and ethical issues; and increase funding for CER.[7]

Luce et al have suggested that the traditional statistical methods (frequentist school) are not suited to CER. Traditional methods require much larger samples because such trials will have a lot more background noise than an RCT and features of the trial might change over time due to information collected during the trial. Bayesian and adaptive analytical models could reduce the sample size but includes previous collected evidence and can adapt to either new interventions in the trial and/or the dropping of less effective interventions.[2]

Volpp and Das[20] note that if we approach CER as simply the comparison of two (or more interventions), what they term Medication A vs Medication B, we might miss the role of individuals' health-related behavior as a major contributing factor. We also must consider the behavior of the health care delivery system.[20]

Garber and Tunis also addressed the concern that CER might ignore individual differences and therefore might be a threat to personalized medicine. They conclude that rather than being a threat, CER might "hasten the discovery of the best approaches to personalization."[21] Presently, they feel we know very little about how a genetic test or genomic medicine might be used to effect in clinical practice and improve health. CER might provide the necessary information for this. Combined with the development of standards for CER and methods that are more rapid, relevant, and efficient, they feel CER could provide a key role in personalized medicine.

One of the challenges for CER is the same as it is for any clinical trial and observation trial: how we determine the quality of the studies. Motheral et al report on a checklist developed to assess retrospective data bases.[22] The checklist contains 27 questions that should be asked, and they cover such things as relevance, reliability, and validity; data lineages; eligibility determination; research design; treatment effects; sample selection; censoring; variable definitions; resource valuation; statistical analysis; generalizability; and data interpretation.

Thorpe et al have published an instrument called Pragmatic Explanatory Continuum Indicator Summary designed specifically to evaluate pragmatic trials.[15] It is based on the assumption that there is a continuum between explanatory and pragmatic trials. The instrument is diagrammatically drawn as a wheel with an "E" at the center, which represents the explanatory end of the continuum. The 10 spokes of the wheel are practitioner adherence, participant compliance, outcomes, follow-up intensity, practitioner expertise (comparison), flexibility of the comparison intervention, practitioner expertise (experimental), flexibility of the experimental intervention, eligibility criteria, and primary analysis. Given any particular pragmatic or explanatory trial, this system can be used to draw a profile on these dimensions. There is still a very subjective element in determining where you would place any study on the individual spokes or dimensions. They also suggest it is a useful tool for designing a trial.

Dreyer developed a system called Good ReseArch for Comparative Effectiveness, or "GRACE Principles." This is a hierarchy of evidence for observational research but with no scoring.[23] The GRACE has three sections, and within each section additional components address the following concerns:

1. Was the plan specified in advance of conducting the study? This should include key research questions, target population, patient characteristics, diseases/conditions, comparators, treatment regimens, and measurements of

effectiveness, safety, tolerability, and intended study size.

2. Was the study conducted, analyzed, and reported in a manner consistent with good practice and reported in enough detail for evaluation and replication? How were the data collected; what checks are there for validity; are data missing; were data compared to similar patients; were alternative explanations for the findings considered and evaluated; was there selection bias, misclassification, detection bias, performance bias, or attrition?

3. How valid is the interpretation of the CE for the population of interest? Many variables can lead to confusion between the treatment and outcome, and we can create a hierarchy here from high-quality evidence where the determinants of treatment are not related to the determinants of outcomes, to middle quality where consistent determinants of treatment are largely unknown, to lower evidence quality, where confounding and bias are likely to be present but little evidence is available.

Dreyer concludes,

> *unless an effect is observed that is much larger than would be expected or larger than could be explained by bias, it is unlikely that the study will contribute meaningfully to clinical decision making. Although there is no unanimity about how large a relative benefit (eg, relative risk of benefitting from a treatment) needs to be in order to be worthy of serious consideration as evidence for decision-making.*[23]

COMPARATIVE EFFECTIVENESS RESEARCH AND COMPLEMENTARY AND ALTERNATIVE MEDICINE: THE REAL CHALLENGE

In many ways, the move to CER should be beneficial for CAM research. Its focus on effectiveness and pragmatic trials is more in keeping with areas of CAM that are not amenable to RCTs without doing considerable "damage" to the very therapy that is being studied by using a reductionist methodology that removes the very strength of a holistic approach. Also, because pragmatic trials allow therapists to do what they wish with regard to treatment—that is, to use their usual and customary care[24]—it is closer to whole-systems research[25] than traditional RCTs.

It also allows for variability in the way individuals are treated in the trial and therefore comes close to "personalized" medicine.[21] This not only solves some ethical issues with regard to care but may assist in recruiting CAM providers to participate.

Due to the fact that it allows the study to focus on those who are getting the care normally and therefore includes populations and subpopulations that are normal in practice, CER leads to results that are more clinically relevant. Also, because they are not explanatory trials, the focus is on what concerns providers and patients: that is, what works (effectiveness) and not what concerns scientists (explanations of why it works, causal analysis, and efficacy). The focus on observational studies also harmonizes well with the CAM field.[11] It places the "P" back in "EBP": that is, it puts the "practice" back into "evidence-based practice."[11] It also provides a role for CAM in EBP based on evidence of effectiveness by broadening the definition of acceptable evidence.

The dilemma is that the choice is between being rigorous and being relevant (clinically). CER research is more clinically relevant, but can it be made more rigorous? RCTs are more rigorous but are frequently not clinically relevant for the populations with whom the provider must deal. As noted earlier, there are systems that attempt to increase the rigor of CERs. Currently, there is a body of work on observation studies that shows they are highly correlated with RCTs[26] and do not exaggerate the effect size if they are good observation studies.[27]

There are, however, some major problems that should concern the CAM community (both providers and researchers) with regard to CER. The first is the threat from once again privileging certain types of evidence. This has been the dominant problem with RCTs, EBP, and CAM and led Holmes, Murray, Perron, and Rail (2006) to state, "the evidence-based movement in health sciences constitutes a good example of microfascism at play in the contemporary scientific arena."[28] Putting aside the hyperbole of the word *fascism*, the article draws attention to the fact that not only are certain forms of evidence privileged in EBP but that privileging evidence is a very social process and a highly political one. EBP does function currently as a powerful ideology and one that is used against CAM. This process of using evidence for ideological purposes against CAM has been documented previously in the CAM literature.[29] The hierarchy of evidence has also been criticized in CAM, and alternative models have been suggested, such the house of evidence[30] or the circular model of evidence as alternative conceptualizations.[31] In a 2008 article in *Lancet,* Rawlins said,

> *Hierarchies place randomized controlled trials (RCTs) at their summit, with various forms of observation studies nestling in the foothills. The notion that evidence can be placed in hierarchies is illusory. . . .Decision makers need to assess and appraise all the available evidence irrespective of whether it has been from randomized controlled trials or observation studies.*[32(p2152)]

A second challenge, however, is that though CER will establish effectiveness (or lack of it), we may be left with no understanding of what makes it effective in any given trial or how we may replicate the intervention. It will also not allow us to separate out any confounding variables or the impact of the encounter as opposed to the therapy. This problem is recognized by those writing about CER. As MacPherson said in 2004, "The pragmatic trial cannot be used to determine precisely what components within the treatment process might have caused any benefits, since it is a package of care that is being evaluated."[24] The only solution to this problem is to collect sufficient information

to determine what was done, but no method for doing this is offered, and no evidence that it can be done rigorously is offered. This is the Achilles heal of the CER, for without this data, we cannot know for sure what was done or whether it can be replicated. All the claims made for the usefulness of CER for making clinical decisions and recommending any form of care become pointless if we cannot replicate the intervention.

One solution is to collect self-reported data from both the provider and the patient, but such data, usually collected retrospectively, is notoriously unreliable. Coulter, Hays, and Danielson (1996)[33] found in their study of preventive care in chiropractic that asking both the provider and the patient independently to indicate what preventive services or recommendations were given immediately after the treatment had a wide variance in responses. Some 25% of the patients failed to indicate services the provider said were provided (which the providers assume reflects the fact patients do not always listen), but more surprisingly, 25% of the patients listed services that the provider did not list.[33] The conclusion from this study was that unless the researcher is observing the actual health encounter itself and recording it directly, no one knows exactly what occurred. Self-reporting is not a solution for this problem.

What is required is ethnographic observation, the type used by anthropologists and sociologists. Chiropractic provides very powerful evidence for the difference between data collected this way and data collected through health services research. It is one of the few CAM groups that has a body of ethnographic studies and a large body of health services research. As the author has noted in earlier papers, the description one gets of the chiropractic role from these two distinct research paradigms is sufficiently different as to suggest that two totally different health care providers are being described.[34] In one description, health services research, the picture of a chiropractor is that of a subspecialist in musculoskeletal problems who uses limited manipulative therapy. In the other, ethnographic research, the chiropractor is depicted as a holistic wellness practitioner who has developed a paradigm around largely musculoskeletal problems but who emphasizes such things as nutrition, weight, diet, exercise, stress, posture, and spirituality.[35]

Rapid Assessment Procedures (RAP)—also called both Rapid Assessment Process and Rapid Evaluation Assessment (REA)—has been widely used as a methodology in the health field.[36,37] Scrimshaw and Hurtado's RAP for studying settings for evaluation uses extensive observation in the sense used by anthropologists and sociologists, including participant observation.[38] REA accomplishes a great deal more in briefer time periods than can many traditional ethnographic projects that rely on a single researcher. Offshoots of REA have been developed such as Rapid Assessment, Response, and Evaluation used by the Office of the Surgeon General for studying HIV/AIDS in metropolitan areas.[39]

REA's strengths are that it is quick and it captures contextual factors: the characteristics of the therapeutic setting, the provider, the patient, and the therapy, including all aspects of the health encounter. This approach is the only entryway into understanding processes at work that are not easily quantified or may be unknown.

A program may be effective, but how and why is it effective? Programs cannot be replicated (or corrected midcourse) unless the elements can be determined. RAP is the method for providing intensive process evaluation on a level not reached by other evaluation models. The weakness is that for the results of RAP to be useful, the researchers must be credible. Therefore, only an experienced and multidisciplinary team can do this type of fast assessment. RAP is not useful for measuring variables that can otherwise be analyzed with quantitative data; therefore, traditional analysis techniques must be used side-by-side with RAP for numeric process and outcome variables.

Rapid Ethnographic Observation has been used previously in studies in CAM. The most extensive use of this method in CAM was by Kelner, Hall, and Coulter (1981) to study the chiropractic health encounter. They conducted ethnographic observation in 72 chiropractic clinics. Two trained observers visited each clinic and spent a minimum of 1 day and a maximum of 2 days observing the clinic, including treatment of patients. From both observations and the quantitative data they collected, the authors were able to create a model of the chiropractic healing encounter consisting of seven stages: (1) the first visit, (2) the formulation of the diagnosis, (3) the explanation to the patient, (4) the negotiation of a plan of treatment, (5) the delivery of the treatment, (6) evaluation of the effects of the treatment, and (7) termination. They were then able to compare and contrast the chiropractic healing encounter with that of biomedicine on these same seven stages and to highlight the unique elements of the chiropractic encounter.[40] But the essential feature of this approach is that it is grounded in what actually occurs in the encounter and captures the full range of contextual features without which it is not possible to fully describe the therapy or to replicate it.

Although the number of clinics observed were less than in the Kelner, Hall, Coulter study, Cowie and Roebuck,[41] Coulehan,[42] and Oths[43] all used ethnographic methods for observing chiropractic clinics to develop a comprehensive account of what occurs in a chiropractic treatment setting.

The bottom line is that if we are to control in any way for confounding variables in CER, we have to be able to measure the confounder[44] or at the very least identify real or potential confounders. Without direct observation, this is very unlikely to happen.

Alternative Approaches

While this author feels that REA/RAP is the most complete method for obtaining the information not generally collected in CER, a mixed-methods approach has been used by Little et al[45] and Hollinghurst et al,[46] so at least some researchers working in CER are adapting methods to enable imputing of the processes being used. Such mixed methods, combining qualitative and quantitative data, have a long history in anthropology and sociology and have been associated with the process of triangulation, a method for validating qualitative data.[47,48]

It could also be argued that standard health services research can accomplish much of this work. Recent articles in *Alternative Therapies in Health and Medicine* by Herman, D'Huyvetter, and Mohler[49] and by Coulter and Khorsan[34] have argued cogently that adopting a health services research paradigm would help resolve some of the issues that the CAM community and those researching CAM face with RCTs.

Last but not least, the field of program evaluation can also claim to do much of this type of work particularly where it includes contextual,[50] formative,[51,52] process,[53] and summative[54] evaluation approaches. Contextual evaluation collects descriptive data used to assess and compare sites. The evaluation focuses on the influence of these factors on the intervention structures, processes, and outcomes. In the initial phases of a program, formative evaluation is used to collect data on intervention components to assess feasibility of proposed activities and to provide recommendations for improving intervention structures and processes. Process evaluation is used to assess the extent to which the intervention components were implemented as planned. Summative evaluation measures the extent to which program goals and objectives were achieved and the intermediate and longer-term impacts of the program. Program evaluation therefore uses both qualitative and quantitative methods allowing triangulation of the data.

CONCLUSION

What has been lacking in the rush to CER is a critical appraisal without which it is likely that the same mistake will be made with CER that has been made with RCTs. When we privilege certain forms of evidence over other forms, we are involved in a process that is not only political in nature but has serious political consequences. The hierarchies of evidence that have been constructed to date have clearly privileged those groups in the health field that have the most of that particular evidence and negatively impacts those groups who do not. This affects the process by legitimizing some groups and not others but also by determining who will be covered in health plans and reimbursed by insurance. As noted earlier, the NIH has not been highly motivated to conduct CER partly because research funded by NIH tends to be explanatory in nature and it has no mechanism to identify the priority questions that might be solved by CER.[7] Its funds go overwhelmingly to biomedicine (70%). Even the $300 million earmarked for AHRQ will not fund a large expansion in its CER studies if AHRQ must meet its other core responsibilities.[7]

What is being proposed in this article can be thought of as an addendum to a recent article by Fønnebø et al,[54] which proposes that the research methods used for drug-related research that begins with the chemical substances and the biologic mechanism and ends with efficacy trials should be reversed in the case of CAM research. The latter begins with clinical practice (context, paradigms, philosophical understanding, and utilization), moves to safety, then to comparative effectiveness, and ends with biological mechanisms. Saying the structure of CAM research should be different is insufficient. CAM also requires the development of rigorous alternative research methods.

However, if all we are doing with CER is privileging a new "trial" model, then we have not advanced very much. Fields advance more effectively through critique. Does the emperor have clothes? Perhaps the emperor has some clothing, certainly more than in the past, but he is certainly not yet fully clothed.

REFERENCES

1. Conway PH, Clancy C. Comparative-effectiveness research—implications of the Federal Coordinating Council's Report. *N Engl J Med*. 2009;361(4):328-330.
2. Luce BR, Kramer JM, Goodman SN, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med*. 2009;151(3):206-209.
3. NCCAM 05-AT-102. Comparative Effectiveness Studies of Complementary and Alternative Medicine. (http://www.grants.gov) Accessed April 27, 2009.
4. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*. 2009;151(3):203-205.
5. Stakeholders Symposium for CER. Center for Medical Technology Policy and the Institute for Integrative Medicine; November 2009; Baltimore, Maryland.
6. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol*. 2009;62(5):499-505.
7. 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(12):1624-1632.
8. Zwarenstein M, Treweek S. What kind of randomized trials do we need? *CMAJ*. 2009;180(10):998-1000.
9. Witt CM. Efficacy, effectiveness, pragmatic trials—guidance on terminology and the advantages of pragmatic trials. *Forsch Komplementmed*. 2009;16(5):292-294.
10. Coulter ID. Observational studies and evidence-based practice: Can't live with them, can't live without them. *J Evid Based Den Pract*. 2003;3(1):1-4.
11. Coulter ID. Putting the practice into evidence-based dentistry. *J Calif Dent Assoc*. 2007;35(1):45-49.
12. Chiappelli F, Cajulis OS, Newman MG. Comparative effectiveness research in evidence-based dental practice. *J Evid Base Dent Pract*. 2009;9(2):57-58.
13. Popper K. Conjectures and refutations. In: Theodore Schick, ed. *Readings in the Philosophy of Science*. London, United Kingdom: Routledge and Kegan Paul; 1963:33-39.
14. Oxman AD, Lombard C, Treweek S, Gagnier JJ, Maclure M, Zwarenstein M. A pragmatic resolution. *J Clin Epidemiol*. 2009;62(5):495-498.
15. Thorpe K, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ*. 2009;180(10):E47-E57.
16. Karanicolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. A new "mechanistic-practical" framework for designing and interpreting randomized trials. *J Clin Epidemiol*. 2009;62(5):479-484.
17. Maclure M. Explaining pragmatic trials to pragmatic policymakers *J Clin Epidemiol*. 2009;62(5):476-478.
18. Karanicolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. The practicalists' response. *J Clin Epidemiol*. 2009;62(5):489-494.
19. Oxman AD, Lombard C, Treweek S, Gagnier JJ, Maclure M, Zwarenstein M. Why we will remain pragmatists: four problems with the impractical mechanistic framework and a better solution. *J Clin Epidemiol*. 2009;62(5):485-488.
20. Volpp KG, Das A. Comparative effectiveness—thinking beyond medication A versus medication B. *N Engl J Med*. 2009;361(4):331-333.
21. Garber AM, Tunis SR. Does comparative-effectiveness research threaten personalized medicine? *N Engl J Med*. 2009;360(19):1925-1927.
22. Motheral B, Brooks J, Clark MA, et al. A checklist for retrospective database studies—report of the ISORP Task Force on Retrospective Databases. *Value Health*. 2003;6(2):90-97.
23. Dreyer NA. GRACE Principles. Good ReseArch for Comparative Effectiveness OBSERVED. *www.graceprinciples.org* Version 25; October 2009:1-10.
24. MacPherson H. Pragmatic clinical trials. *Complement Ther Med*. 2004;12(2-3):136-140.
25. Jonas WB, Beckner W, Coulter I. Proposal for an integrated evaluation model for the study of whole systems health care in cancer. *Integr Cancer Ther*. 2006;5(4):315-319.
26. Benson K, Hartz A. A comparison of observational and randomized, controlled trials. *New Eng J Med*. 2000;342(25):1878-1886.
27. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New Eng J Med*. 2000;342(25):1887-1892.
28. Holmes D, Murray SJ, Perron A, Rail G. Deconstructing the evidence for evidence-based discourse in health sciences: truth, power and fascism. *Int J Evid Based Health*. 2006;4(3):180-186.
29. Coulter ID. Sociological studies of the role of the chiropractor: an exercise in ideological hegemony? *J Manipulative Physiol Ther*. 1991;14(1):51-58.
30. Jonas WB. The evidence house: how to build an inclusive base for complementary medicine. *West J Med*. 2001;175(2):79-80.
31. Walach H, Falkenberg T, Fønnebø V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol*. 2006 Jun 24;6:29.
32. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet*. 2008;372(9656):2152-2161.
33. Coulter ID, Hays RD, Danielson CD. The role of the chiropractor in the changing health care system: from marginal to mainstream. *Res Sociol Health Care*. 1996;13A:95-117.
34. Coulter ID, Khorsan R. Is health services research the Holy Grail of CAM research?

Altern Ther Health Med. 2008;14(4):40-45.
35. Coulter ID. Competing views of chiropractic: health services research versus ethnographic observation. In: Oths KS, Hinojosa SZ, eds. *Healing by Hand: Manual Medicine and Bonesetting in Global Perspective.* Walnut Creek, CA: AltaMira Press; 2004:43-62.
36. Scrimshaw S, Hurtado E. *RAP: Rapid Assessment Procedures For Nutrition and Primary Care. Anthropological Approaches to Improving Programme Effectiveness*. Vol 11. Los Angeles, CA: UCLA; 1987.
37. Beebe J. *Rapid Assessment Process: An Introduction*. Walnut Creek, CA: AltaMira; 2001.
38. Scrimshaw SC, Hurtado E. Anthropological involvement in the Central American diarrheal disease control project. *Soc Sci Med.* 1988:27(1):97-105.
39. Trotter RT, Needle RH, Goosby E, et al. A methodological model for rapid assessment, response, and evaluation: The RARE Program in Public Health. *Field Methods.* 2001;13(2):137-159.
40. Kelner M, Hall O, Coulter I. *Chiropractors: Do They Help?* Toronto, Ontario: FitzhenryWhitesides; 1980.
41. Cowie JB, Roebuck JB. *An Ethnography of a Chiropractic Clinic: Definitions of a Deviant Situation.* New York, NY: Free Press; 1975.
42. Coulehan JL. Chiropractic and the clinical art. *Soc Sci Med.* 1985;21(4):383-390.
43. Oths K. Communication in a chiropractic clinic: how a D.C. treats his patients. *Cult Med Psychiatry.* 1994;18(1):83-113.
44. Strom BL. Methodologic challenges to studying patient safety and comparative effectiveness. *Med Care.* 2007;45(10 Suppl 2):S13-S15.
45. Little P, Lewith G, Webley F, et al. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain. *Br J Sports Med.* 2008;42(12):965-968.
46. Hollinghurst S, Sharp D, Ballard K, et al. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain: economic evaluation. *BMJ.* 2008 Dec 11;337:a2656.
47. Jick TD. Mixing qualitative and quantitative methods: triangulation in action. *Administration Sci Q.* 1979;24(4): 602-611.
48. de Munck VC, Sobo EJ, eds. *Using Methods in the Field: A Practical Introduction and Casebook*. Walnut Creek, CA: AltaMira Press; 1998.
49. Herman PM, D'Huyvetter K, Mohler MJ. Are health services research methods a match for CAM? *Altern Ther Health Med.* 2006;12(3):78-83.
50. Israel BA, Cummings KM, Dignan MB, et al. Evaluation of health education programs: current assessment and future directions. *Health Educ Q.* 1995;22(3):364-389.
51. Brindis C, Hughes DC, Halfon N, Newacheck PW. The use of formative evaluation to assess integrated services for children. The Robert Wood Johnson Foundation Child Health Initiative. *Eval Health Prof.* 1998;21(1):66-90.
52. Scheirer MA. Designing and using process evaluation. In: Hatry HP, Wholey JS, Newcomer KE, eds. *Handbook of Practical Program Evaluation*. San Francisco, CA: Jossey-Bass; 1994:40-68.
53. Rossi PH, Freeman HE, Lipsey MW. *Evaluation: A Systematic Approach.* 6th ed. Thousand Oaks, CA: Sage Publications; 1999.
54. Fønnebø V, Grimsgaard S, Walach H, et al. Researching complementary and alternative treatments—the gatekeepers are not at home. *BMC Med Res Methodol.* 2007 Feb 11;7:7.