Complementary Therapies in Medicine (2005) 13, 206–212

ELSEVIER

Complementary Therapies in Medicine

www.elsevierhealth.com/journals/ctim

Complementary and alternative medicine whole systems research: Beyond identification of inadequacies of the RCT

Marja J. Verhoef[a,*], George Lewith[b], Cheryl Ritenbaugh[c], Heather Boon[d], Susan Fleishman[e], Anne Leis[f]

[a] *Department of Community Health Sciences, University of Calgary, Faculty of Medicine, Canada*
[b] *Complementary Medicine Research Unit, Primary Medical Care, Aldermoor Health Centre, Southampton, UK*
[c] *Department of Family and Community Medicine, University of Arizona, AZ, USA*
[d] *Faculty of Pharmacy, University of Toronto, Ont., Canada*
[e] *Integrative Medicine Resource Group, Tucson, AZ, USA*
[f] *Department of Community Health and Epidemiology, University of Saskatchewan, Sask., Canada*

Available online 15 August 2005

Summary Complementary and alternative medicine (CAM) often consists of whole systems of care (such as naturopathic medicine or traditional Chinese medicine (TCM)) that combine a wide range of modalities to provide individualised treatment. The complexity of these interventions and their potential synergistic effect requires innovative evaluative approaches. Model validity, which encompasses the need for research to adequately address the unique healing theory and therapeutic context of the intervention, is central to whole systems research (WSR). Classical randomised controlled trials (RCTs) are limited in their ability to address this need. Therefore, we propose a mixed methods approach that includes a range of relevant and holistic outcome measures. As the individual components of most whole systems are inseparable, complementary and synergistic, WSR must not focus only on the ''active'' ingredients of a system. An emerging WSR framework must be non-hierarchical, cyclical, flexible and adaptive, as knowledge creation is continuous, evolutionary and necessitates a continuous interplay between research methods and ''phases'' of knowledge. Finally, WSR must hold qualitative and quantitative research methods in equal esteem to realize their unique research contribution. Whole systems are complex and therefore no one method can adequately capture the meaning, process and outcomes of these interventions.

* Corresponding author. Tel.: +1 403 220 7813; fax: +1 403 270 7307.
E-mail address: mverhoef@ucalgary.ca (M.J. Verhoef).

0965-2299/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ctim.2005.05.001

Complementary and alternative medicine (CAM) often consists of whole systems or disciplines of health care (such as naturopathic medicine, traditional Chinese medicine and ayurveda) that include a wide range of modalities—from diet and herbal products to acupuncture and yoga. Most of the research to date has focused on the effectiveness and efficacy of individual modalities rather than on the systems in which they are found.

Due to the complexity of whole systems, study designs and methods need to acknowledge the uniqueness of such systems as complex and intrinsically adaptive. *Whole systems* can be defined as ''approaches to health care in which practitioners apply bodies of knowledge and associated practices in order to maximize the patients' capacity to achieve mental and physical balance and restore their own health, using individualised, non-reductionist approaches to diagnosis and treatment''.[1] Furthermore, we suggest that in whole systems the practitioner—patient relationship and therapeutic environment may play a central role and that this role continues to evolve over time.[1]

Whole systems research (WSR) is an emerging research framework specific for the investigation of the effectiveness of whole systems of health care. The aim of WSR is to use appropriate research designs and methods so that all aspects of any internally consistent approach to treatment, or a whole system, can be assessed within its unique explanatory model.[1] This in turn allows for the assessment of the explanatory model itself. Reflecting on the definition of whole systems, it is evident that WSR must acknowledge an individualised, patient centred and participatory approach to diagnosis and treatment and a process of healing that collaboratively combines patient and practitioner knowledge and skills, thus enhancing healing.[2] In this paper, we discuss design strategies that may be used to evaluate whole systems and recommend directions for further methodological development.

Why whole systems research?

Assessing the efficacy of whole systems is more complex than assessing the efficacy of single modalities or treatments. The fundamental challenge is to acknowledge all key components of the intervention as a network, consequently none of the components can be considered in isolation. Research methods that focus on the identification of ''active'' ingredients of an intervention through reductionist strategies are rarely appropriate for whole systems.[3]

Classical randomised controlled trials (RCTs) are widely accepted as the gold standard for answering questions of efficacy, but have limitations, especially when applied to the study of CAM whole systems. They may have powerful internal validity, but poor external validity (generalisability) depending on the specific randomisation and sampling procedures used. The standardised diagnostic processes required in RCTs are not always feasible or appropriate within CAM whole systems that rely on individualised diagnostic principles unique to different healing systems. Furthermore, the individualised treatment packages provided in the clinical practice of many whole systems contradict the necessity for standardised interventions to be delivered within an RCT design. Randomisation may also pose a problem, as many whole systems depend on the willingness and readiness of the patient to participate in the healing process. Patient perceptions and expectations about the therapeutic relationship may be intertwined with treatment effects. The process of an RCT may therefore compromise this therapeutic process and its outcome.[4] WSR needs to be comprehensive and include multiple systematic strategies that are tailored to the philosophical assumptions of a particular system under investigation while maintaining rigour and relevance.

Model validity: a unique consideration for whole systems research

A key component of WSR is the requirement for model validity. Model validity encompasses the need for the research to adequately address the unique healing theory and therapeutic context of the intervention that is being assessed.[5] Cassidy[6] describes model validity as *paradigm fit* and suggests that a paradigm fit exists if the explanatory model of the system under investigation and the research design are aligned. This requirement is not unique to CAM or to whole systems, but must be considered when studying healing systems outside the biomedical paradigm. Criteria to assess model validity[5] include:

1. *Representativeness*—to assess whether the intervention was consistent with current practice, likely to produce an effect and clearly and adequately described.
2. *Equipoise and credibility*—to assess equipoise on the part of both patients and practitioners with respect to the intervention being evaluated, the credibility of the intervention and expected treatment effects.

3. *Model congruity*—to assess whether the diagnosis, intervention and outcomes fit the system under investigation.
4. *Context*—to assess patient and practitioner confidence in the intervention and whether the intervention was sensitive to the culture, family and meaning that the patient ascribes to an intervention.

Internal, external and model validity are not independent of each other. In designing WSR, however, attention to model validity in addition to internal and external validity is crucial, as the philosophical assumptions of the CAM whole system under investigation often differ from the classical (biomedical) methods.

An example of a simple technique that, when applied, can enhance model validity is *double classification,* or classifying study participants on both a conventional medicine diagnosis and a diagnosis from the whole system that is being evaluated. Research designs that ignore the diagnostic approach of a CAM whole system have weak model validity and can only achieve a weak test of the system's potential. For example, Collet et al. (personal communication, April 5, 2005) are studying the effectiveness of traditional Chinese medicine (TCM) in improving quality of life in cancer patients. The individualised treatment patients receive is based on the TCM diagnosis. This process was first described by Fisher et al. in the context of homeopathic trials when a conventional diagnosis was used first, followed by a classical homeopathic diagnosis to generate individualised treatment.[7]

Design strategies to evaluate whole systems

Approaches to evaluate CAM whole systems may be described as variations of the RCT design, such as the adaptations suggested by Ernst.[8] The Institute of Medicine[9] similarly suggests adaptations of RCTs to assess the effectiveness of CAM, but also includes other study designs, such as observational studies. Ernst's suggestions, however, are limited to the assessment of single CAM modalities, and while the Institute of Medicine recognizes that many CAM approaches are delivered as whole systems, or ''bundles'' of therapies, neither identifies the need for strategies to address model validity. The suggested study designs also have limited capability to capture the process and context that are so crucial to the effectiveness of CAM whole systems. We recommend combining qualitative and quantitative research (''mixed methods'') as an optimal approach to WSR.

Variations of the classical RCT design

Pragmatic trials allow for the assessment of individualised treatment approaches, which are fundamental for many CAM whole systems. Control is maintained by randomising participants to treatment groups, but the intervention is intended to represent ''real world'' care, and thus to enhance external validity. Model validity may also be enhanced if adequate attention is paid to appropriate outcome measures. For example, Vickers et al.[10] conducted a large, pragmatic RCT to compare a policy of ''use acupuncture'' and ''avoid acupuncture'' in general practice for chronic headache. All treatments were individualised and a range of outcome measures were assessed, including headache frequency and severity, medication use, health status, days off sick and use of health resources. A two-arm pragmatic trial, however, does not allow for the evaluation of the relative effectiveness of various components of the intervention.

Factorial designs compare single modalities to a combination of modalities to allow for the assessment of multiple interaction effects between treatments. For example, the UK back pain exercise and manipulation (BEAM) trial evaluated the effectiveness of adding: (1) an exercise program; (2) manipulation; or (3) manipulation followed by exercise, to ''best care'' in the treatment of back pain in general practice.[11] Through the comparison of multiple treatment groups, a factorial design can provide insight into the relative effectiveness of various components of an intervention as well as the synergistic effects two (or more) components may have. Combining pragmatic and factorial designs can enhance external validity and potentially also internal and model validity.

Preference trials are useful to address the argument that in many CAM modalities or whole systems, patient preference is correlated with treatment effectiveness and therefore randomisation is not appropriate. In preference trials, participants with no treatment preferences are randomised as usual but those with preferences receive their preferred treatment, allowing for the assessment of the interaction between treatment preference and treatment outcome.[12]

n-of-1 Trials address the critique that the classical RCT design does not provide information about individuals, but only an average effect of a treatment on a group of individuals. The n-of-1 design is a single-patient trial with multiple crossovers

between a treatment period and a placebo or standard treatment period.[13] To enhance internal, external and model validity, patient preference may be taken into consideration when determining an appropriate intervention and the intervention may be individualised. n-of-1 Trials are limited, however, as they only provide individual level data,[8] making applications to groups or populations difficult. Combining the results of separate n-of-1 trials may be appropriate to address this concern.[9]

Observational studies

Recent reviews provide strong evidence that well-designed observational studies yield results comparable to RCTs,[14,15] and may be less complex and expensive to operate. Observational studies appear to have strong external validity as individualised treatments may be applied and non-experimental practice may be evaluated.[16] A wide range of treatment outcomes may also be assessed, which, if chosen appropriately, have the potential to enhance model validity.

When used for the assessment of whole systems, observational studies should 'take a long-term perspective',[17,18] in comparison to RCTs, given the often slow and progressive relief of symptoms and/or the initial deterioration of symptoms in CAM. Observational studies, however, have no comparison or control group and positive results may simply reflect regression to the mean. These concerns may be addressed through various strategies to enhance internal validity, such as thorough medical documentation, including independent chart reviews and predicted survival for each patient,[17] baseline measurement, or a (non-randomised) comparison group,[18] which could take the form of a matched group or waiting list control or comparison to population norms.[16] Further validity could be provided by combining RCTs with an observational arm, thus creating (at least) a three-arm study.

Mixed methods research

Combining both quantitative and qualitative research methods has most potential to effectively evaluate whole systems of health care. Modified RCTs and well-conducted observational studies are paramount to assess effectiveness, but it is not enough to gain evidence from such research designs alone. Whole systems of health care are complex, and the structure and process of the approach to healing may be as important to assess as the intervention itself. Qualitative methods provide the opportunity to explore the meaning that patients ascribe to an intervention or system, the process and context by which healing occurs, outcomes that are relevant and meaningful to patients (and sometimes unexpected by practitioners and researchers), how an intervention fits with a participant's life, and the role that expectations may play in healing.[19] Qualitative methods may also assist in generating new hypotheses, and informing the conceptual link between treatment and outcome, thus enhancing model validity.[20] One particularly useful approach to combining qualitative and quantitative methods is nesting qualitative studies within rigorous adaptations of the RCT.

Ideally each approach should be held in equal esteem.[21] In clinical trials implemented at the Oregon Centre of Complementary and Alternative Medicine (OCCAM), qualitative and quantitative results are intertwined. Across participants, responses to pain-scale items and descriptions of pain obtained during individual interviews are compared to explore what is meant by a pain rating of ''4 on a scale of 0—10''. Within participant cases, qualitative and quantitative data are combined to explore how an individual's social, cultural, medical history and psychosocial characteristics influence their healing experience.

When qualitative and quantitative methods are appropriately combined, a comprehensive research design may result that addresses the complex nature of a whole system that is patient centred, relevant and sensitive to the philosophical assumptions of that system.[3]

Appropriate outcome measures

It is important that expected outcomes of whole system interventions are clinically important from both the patient and practitioner perspective, are measured at a time appropriate to see an effect[22] and are relevant to the intent and philosophy of the intervention. Individualisation of treatment implies the use of individual endpoints to indicate successful treatment. Instruments such as Goal Attainment Scaling (GAS)[23] and Measure Your Medical Outcome Profile (MYMOP)[22] are participatory approaches that are valid and individualised and fit within a whole systems philosophy. Proposals to develop valid measures for other concepts relevant to whole person healing, such as transformation and 'unstuckness', are also in progress (Bell, Ritenbaugh, Verhoef et al.).

CAM whole systems affect more than one aspect of a patient's life and maximize the patient's capacity to achieve mental and physical balance and restore his/her own health.[1] Qualitative research participants have used terms

such as ''awareness'', ''openness'', ''peace'', ''transformation'', ''reflection'' and ''energy[24]'' to express the effects of healing, in addition to standardised quality of life measures. To appropriately evaluate whole systems, expected outcomes must adequately reflect the variety of ways in which a patient may benefit from an intervention. The use of multiple outcome measures (both objective and subjective) assists in achieving internal, external and model validity and developing hypotheses about the mechanisms and potential of whole systems.

Miller et al.[25] have identified six different domains of outcomes, including (1) changes in physiological parameters (e.g., CD4, PSA); (2) symptom resolution; (3) cure (e.g., biomarkers, disease outcomes); (4) improved sense of well-being; (5) movement towards wholeness (e.g., sense of coherence, transformation); and (6) enhanced relationships (e.g., work satisfaction, social support). While measures in most of these domains are currently available, there are no measures that assess healing systems as a whole.[26] If each domain is assessed separately, the impact that the whole system may have over and above its components is lost. Miller et al.[25] suggest that ''one could conceivably construct a composite measure of healing using the different outcomes domains.'' In order to do this, it is important that a conceptual framework of the multiple effects of whole systems of healing be developed.[22,26] Such a framework would be philosophically sensitive to the whole system and would address all components within the system, including the structure, process and context in which the intervention is delivered.

Moving forward

Several strategies to evaluate CAM whole systems have been presented; however, the question still remains—how do we know if whole systems are effective? While individualised data are necessary, a means to interpret these data as a whole is clearly needed in order to strengthen the evidence required for uptake in practice. The complex and individualised nature of whole systems suggests that one well-designed, multi-method study is not sufficient to indicate effectiveness. A determination of effectiveness will best be approached through a program of research, as opposed to individual research projects. The necessity for a program of research was recognized by members of the UK Medical Research Council's (MRC) Health Services and Public Health Research Board, who developed a sequential framework for the evaluation of complex interventions for conventional medicine.[27] This framework is a reasonable starting point for the evaluation of whole systems and identifies the importance of using a range of research methodologies underpinned with a clear hypothesis. It does not, however, explicitly address the unique philosophical assumptions that underlie whole systems, allow for individualised treatment packages to be applied or assessed, or allow for the evaluation or description of the context in which an intervention is delivered. It focuses on a far more reductionist approach that fails to address the potentially synergistic process within whole systems.

The Institute of Medicine[9] suggests that the same principles and standards should apply to CAM treatments as currently apply to conventional medical treatments, and that common methods, measures and standards for the generation and interpretation of evidence should be developed. Their stated logic is that in all effectiveness research, the hypothesis to be tested is that ''Treatment A produces Health Benefit Y''. However, the hypothesis that ''Treatment A in the presence of *x*, *y* and *z* produces Health Benefit Y''[28] is more precise as it identifies the impact of contextual variables and, thus, appears more appropriate with respect to complex systems. As central themes within WSR research we need to address model validity, the need for relevant and individualised patient centred outcomes, the patient—practitioner relationship, the context and environment of the intervention and the interactions between the various components of the system. Further, to respect that whole systems do not necessarily work in a linear fashion, more frequent data collection (both qualitative and quantitative) over a longer period of time may be needed. In addition, innovative statistical methodology will undoubtedly be required to examine the large volume of individualised data that such research generates.

Given the imperative to evaluate CAM whole systems, three fundamental components of an emergent WSR framework should be considered. First, the individual components of most whole systems are inseparable, complementary and synergistic and therefore WSR must not focus only on the elements that may initially be perceived as the ''active'' ingredients of a system, such as homeopathic medication in the context of a homeopathic treatment. It is essential to understand what the components of the system are and how they may interact and influence outcomes, but initially the focus cannot be on any one component of the system that may appear to have ''the most'' benefit. Second, we must recognize that creating knowledge is a continuous and evolutionary process. A WSR

framework must not suggest progression through mutually exclusive stages to provide one definitive answer regarding the effectiveness of that system once all stages have been completed. There must be a continuous interplay between a variety of research methods and a variety of ''phases'' of knowledge, so that assumptions may continuously be revisited in light of new knowledge. Therefore, a WSR framework must be non-hierarchical, networked, cyclical, flexible and adaptive. Finally, a framework must hold qualitative and quantitative research methods in equal esteem and realize the unique contribution that each may make. Whole systems are complex and therefore no one method can adequately capture the meaning, process and outcomes of these interventions.

Conclusion

How best to study whole systems of health care is an issue that is not unique to CAM, it also applies to complex interventions in conventional health care, such as multidisciplinary chronic care, patient-centred primary care, psychotherapy, and palliative care. While WSR is potentially very costly, its advancement will assist in understanding the role of all components of the effectiveness of any intervention, including sensitivity to the philosophical assumptions on which an intervention is based, and thus, will be relevant to both CAM and conventional medicine. Whereas the RCT is the gold standard to test causal relationships in biomedicine, no well-developed similar standard, taking into account complex interrelationships, exists to guide WSR. One of the most difficult aspects of a whole system is the recognition of its non-linear nature, which does not lend itself to a linear cause and effect model. While important progress has been made in identifying unique methodological approaches for WSR, more work is needed to develop valid and powerful designs (including outcomes measures and analytic strategies) that transforms and enhances current clinical research methodology so it can become increasingly relevant to complex interventions.

Given that WSR needs to take into account the diverse philosophical and conceptual views that underpin whole systems and must incorporate a variety of designs and innovative analytic techniques, multidisciplinary teams are essential.[1] The recent development of CAM research networks across the world has the potential to support this vision and to create collaborative teams to ensure is a consolidated effort to broaden and deepen CAM whole systems methodology.

References

1. Ritenbaugh C, Verhoef M, Fleishman S, Boon H, Leis A. Whole systems research: a discipline for studying complementary and alternative medicine. *Altern Ther Health Med* 2003;**9**:32—6.
2. Boon H, Verhoef M, O'Hara D, Findlay B, Majid N. Integrative healthcare: arriving at a working definition. *Altern Ther Health Med* 2004;**10**:48—56.
3. Carter B. Methodological issues and complementary therapies: researching intangibles? *Complement Ther Nurs Midwifery* 2003;**9**:133—9.
4. Weatherley-Jones E, Thompson EA, Thomas KJ. The placebo-controlled trial as a test of complementary and alternative medicine: observations from research experience of individualised homeopathic treatment. *Homeopathy* 2004;**93**:186—9.
5. Lewith G, Walach H, Jonas WB. Balanced research strategies for complementary and alternative medicine. In: Lewith G, Walach H, Jonas WB, editors. *Clinical research in complementary therapies: principles, problems and solutions*. Edinburgh: Churchill Livingstone; 2002. p. 1—27.
6. Cassidy CM. Unraveling the ball of string: reality, paradigms, and the study of alternative medicine. Via the Internet (http://www.healthy.net/LIBRARY/Articles/Advances/CASSIDY.htm). Accessed March 2004.
7. Fisher P, Greenwood A, Huskisson EC, Turner P, Belon P. Effect of homoeopathic treatment on fibrositis (primary fibromyalgia). *BMJ* 1989;**299**:365—6.
8. Ernst E. RCTs for CAM. *Focus Altern Complement Ther* 2005;**10**:9—12.
9. Institute of Medicine. *Complementary and alternative medicine in the United States. Board on population health and public health practice. Chapter 4. Need for innovative designs in research on CAM and conventional medicine*. The National Academic Press; 2005. p. 108—28.
10. Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, et al. Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial. *BMJ* 2004;**328**:744.
11. UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *BMJ* 2004:329.
12. Bower P, King M, Nazareth I, Lampe F, Sibbald B. Patient preferences in randomised controlled trials: conceptual framework and implications for research. *Soc Sci Med* 2005;**61**:685—95.
13. Johnston BC, Mills E. n-of-1 randomized controlled trials: an opportunity for complementary and alternative medicine evaluation. *J Altern Complement Med* 2005;**10**:979—84.
14. Benson K, Hartz AA. Comparison of observational studies and randomized controlled trials. *N Engl J Med* 2000;**342**:1878—86.
15. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887—92.
16. Thomas K, Fitter M. Possible research strategies for evaluating CAM interventions. In: Lewith G, Walach H, Jonas WB, editors. *Clinical research in complementary therapies: principles, problems and solutions*. Edinburgh: Churchill Livingstone; 2002. p. 59—91.
17. Cunningham AJ. A new approach to testing the effects of group psychological therapy on length of life in patients with metastatic cancers. *Adv Mind Body Med* 2002;**18**:5—9.

18. Long AF, Mercer G. Challenges in researching the effectiveness of complementary therapies. *J Contemp Health* 1999;**8**:13—9.
19. Verhoef MJ, Casebeer AL, Hilsden RJ. Assessing efficacy of complementary medicine: adding qualitative research methods to the ''gold standard''. *J Altern Complement Med* 2002;**8**:275—81.
20. Sandelowski M. Focus on qualitative methods: using qualitative methods in intervention studies. *Res Nurs Health* 1996;**19**:359—64.
21. Vuckovic N. Integrating qualitative methods in randomised controlled trials: the experience of the Oregon Center for Complementary and Alternative Medicine. *J Altern Complement Med* 2002;**8**:225—7.
22. Long AF. Outcome measurement in complementary and alternative medicine: unpicking the effects. *J Altern Complement Med* 2002;**8**:777—86.
23. Becker H, Stuifbergen A, Rogers S, Timmerman G. Goal attainment scaling to measure individual change in intervention studies. *Nurs Res* 2000;**49**:176—80.
24. Verhoef M. Placebo in qualitative studies—where is it hiding? Presented at Research Seminar on Complementary and Alternative Treatment, Sommarøy, 15 March 2005.
25. Miller WL, Crabtree BF, Duffy B, Epstein RM, Stange KC. Research guidelines for assessing the impact of healing relationships in clinical medicine. *Altern Ther Health Med* 2003;**9**:A80—95.
26. Quinn JF, Smith M, Ritenbaugh C, Swanson K, Watson MJ. Research guidelines for assessing the impact of healing relationships in clinical nursing. *Altern Ther Health Med* 2003;**9**:A65—79.
27. Medical Research Council, A framework for development and evaluation of RCTs for complex interventions to improve health. Via the Internet (http://www.mrc.ac.uk/pru/pdf-mrc_cpr.pdf). Accessed March 2004.
28. Temoshok LR. Rethinking research on psychosocial interventions in biopsychosocial oncology: an essay written in honour of the scholarly contributions of Bernard H. Fox. *Psychooncology* 2004;**13**:460—7.