EXPERT OPINIONS ON METHODOLOGY: DEVELOPMENT OF CANCER CAM SYMPTOM RESEARCH

Final Report from the Expert Panel on Symptom Research

Bethesda, Maryland
November 29–30, 2001

TABLE OF CONTENTS

Preface.....v

Expert Panel on Symptom Research: Panel Members..... vii

Executive Summary

Introduction.....1

The Placebo and Complementary and Alternative Medicine:
Historical, Scientific, and Methodological Considerations 2

Complementary and Alternative Medicine (CAM) Clinical Trials for
Symptoms: Development of Appropriate Interventions.....5

Assessment of Cancer-Related Symptoms: Relevance for CAM
Research Methodology 7

Theory, Hypothesis Development, and Randomized Clinical Trials in
Complementary and Alternative Medicine; A Beginning at Separating
Science from Political Science.....9

Development of Botanical "New Drugs": FDA Regulations 10

Ethical Challenges in Complementary and Alternative Medicine
Cancer Symptom Research.....12

Statistical Issues in the Study of CAM Interventions 15

Strategies for Grant Preparation in Cancer CAM Symptom Research.....19

References.....21

Original Papers by Panelists 23

PREFACE

As cancer patients continue to explore alternative treatments and practices, the need for reliable scientific data increases. Consumers believe that if a practice or product has been in use for hundreds of years, it must be effective. Many also believe that if a product is "natural," it must be safe. Unfortunately, many of these assumptions are not based upon scientific evidence. The use of complementary and alternative medicine (CAM) may not only expose patients to potential toxicities, it may also compromise the effectiveness of conventional treatments. In addition, there may be therapeutic interventions of great value in the treatment of cancer and its symptoms that have yet to be rigorously evaluated.

The National Cancer Institute (NCI) remains devoted to the rigorous investigation of any potential treatments and modalities in the prevention and treatment of cancer and its symptoms regardless of its unconventional or unexpected source. In 1998, the National Cancer Institute established the Office of Cancer Complementary and Alternative Medicine (OCCAM) to support the scientifically rigorous study of CAM modalities as they relate to the diagnosis, prevention, and treatment of cancer and its symptoms. The NCI's OCCAM is committed to developing the foundation for scientifically rigorous research in cancer CAM and symptom management.

NCI Expert Panels:

Whether one is reviewing the literature to guide clinical practice or interested in conducting research in CAM, it is critically important to understand the unique challenges within CAM research methodology. As the field of CAM research has developed, the need for well-developed research methodologies has become apparent.

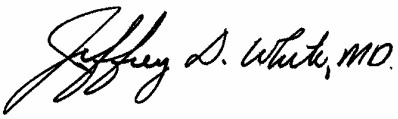
In recognition of this need, the NCI's Office of Cancer Complementary and Alternative Medicine established a series of expert panels to assess and critique the state of the science in research methodologies in CAM cancer research. Panelists from both conventional and CAM research will apply their knowledge and expertise to specific topic areas within cancer CAM. Panelists are to identify the major methodological challenges in cancer CAM research and propose potential solutions. It is our expectation that this process will assist grant applicants by illustrating the types of issues that should be addressed in cancer CAM research proposals.

The first panel focused on cancer symptom research. The discussion and conclusions raised by this panel are focused primarily on pain research but are applicable to other symptoms as well.

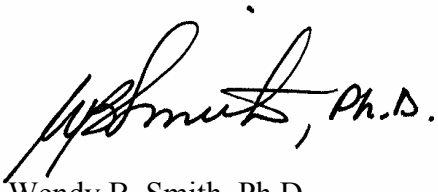
Expert Opinions on Methodology: Development of Cancer Symptom Research

Some of the most critical topics in CAM research methodology were identified and the NCI commissioned experts in those fields to write and present papers at a meeting of the panel. Other experts were invited to serve on the panel and comment on the papers and lead discussion among all the experts. Panelists, many of whom have been successful NIH grant applicants and have served on NIH review committees, identified research methodology challenges within these topics and proposed potential solutions and strategies to address these challenges.

It is our hope that investigators will find this document useful in the development of their research programs and in the preparation of grant proposals. We would like to express our appreciation to the distinguished members of this panel for their thoughtful and valuable contribution to the development of this field.



Jeffrey D. White, MD
Director
Office of Cancer Complementary and Alternative Medicine
National Cancer Institute



Wendy B. Smith, Ph.D.
Program Director, Research Development & Support Program
Office of Cancer Complementary and Alternative Medicine
National Cancer Institute

EXPERT PANEL ON SYMPTOM RESEARCH PANEL MEMBERS

Co-Chairpersons

Wendy B. Smith, PhD
Program Director
Office of Cancer Complementary and Alternative Medicine
National Cancer Institute
Bethesda, MD

Kristine A. Nelson, MD
Director
Palliative Medicine Research Program
The Cleveland Clinic Foundation
Cleveland, OH

Participants

Brian Berman, MD
Professor of Family Medicine
Director, Complementary Medicine Program
Kerman Hospital
University of Maryland
Baltimore, MD

Shaw T. Chen, MD, PhD
Associate Director
Center for Drug Evaluation and Research
Food and Drug Administration
Rockville, MD

Charles S. Cleeland, PhD
McCullough Professor of Cancer Research
Director, Pain Research Group
University of Texas
M.D. Anderson Cancer Institute
Houston, TX

Richard J. Gralla, MD
Professor of Medicine
Herbert Irving Comprehensive Cancer Center
Columbia-Presbyterian Medical Center
New York, NY

Ted Kaptchuk, OMD
Associate Professor of Medicine
Harvard Medical School
Associate Director, Center for Alternative Medicine Research and Education
Beth Israel Deaconess Medical Center
Boston, MA

Patricia A. McGrath, PhD
Professor of Pediatrics and Psychologist
Hospital for Sick Children Toronto
The University of Western Ontario
London, Ontario, Canada

Christine Miaskowski, RN, PhD, FAAN
Professor and Chair, Department of Physiological Nursing
University of California San Francisco
San Francisco, CA

Gary R. Morrow, PhD
Professor of Radiation Oncology and Psychiatry
University of Rochester Cancer Center
Rochester, NY

Jeff Sloan, PhD
Lead Statistician
Cancer Center Statistics Department
Mayo Clinic
Rochester, MN

Jacquelyn Slomka, PhD
Assistant Professor
University of Texas Houston School of Public Health
Houston, TX

Andrew Vickers, DPhil
Assistant Attending Research Methodologist
Memorial Sloan-Kettering Cancer Center
New York, NY

Lonnie K. Zeltzer, MD
Director
Pediatric Pain Program
UCLA School of Medicine
Los Angeles, CA

EXECUTIVE SUMMARY

INTRODUCTION

CAM expert panel members, left to right, back row: Shaw Chen, Lonnie Zeltzer, Charles Cleeland, Jacquelyn Slomka, Chistine Miaskowski, Gary Morrow, Jeff Sloan, Andrew Vickers, Brian Berman, Ted Kaptchuk; left to right, front row: Jeffrey White, Kristine Nelson, Wendy Smith, Richard Gralla, Patricia McGrath.

The Office of Cancer Complementary and Alternative Medicine (OCCAM), within the National Cancer Institute (NCI), was established in October, 1998 to coordinate and enhance the activities of the NCI in the field of Complementary and Alternative Medicine (CAM). The responsibilities of OCCAM include the coordination of research and information initiatives across the NCI that are focused on CAM as it relates to the diagnosis, prevention, and treatment of cancer, cancer-related symptoms, and side-effects of conventional treatment.

CAM research often involves novel concepts and claims, and uses complex systems of practice that need systematic, explicit, and comprehensive knowledge and skills to investigate. Preparation of competitive grant applications in CAM research may be particularly challenging. NCI provides most research funds through the grant application process and review of such grants has identified a number of major methodological roadblocks that ultimately hinder potential advancement in CAM research. To this end, the OCCAM convened an Expert Panel on the State of the Science Methodologies in CAM Cancer Research to identify and develop standard methodologies for CAM cancer symptom research.

THE PLACEBO AND COMPLEMENTARY AND ALTERNATIVE MEDICINE: HISTORICAL, SCIENTIFIC, AND METHODOLOGICAL CONSIDERATIONS

Ted J. Kaptchuk, OMD, Harvard Medical School, author

Brian M. Berman, MD, University of Maryland School of Medicine, author and discussant

Although the placebo controls are routinely used for evaluation of mainstream pharmacotherapy, they were first used to test unconventional therapies over 150 years ago (Kaptchuk 1998). Since then, the limited use of placebo controls in CAM research and the questionable placebo effects of CAM have been central to the division between mainstream therapy and unconventional medicine (Kaptchuk 2001). One reason for the limited use of placebo controls in CAM research is based on the complexity of developing an appropriate placebo for alternative interventions that are not based on a single or readily identifiable chemical entity or procedure.

Are Placebo Controls the Only Legitimate Method for Establishing Efficacy?

It is well recognized that placebo-controlled, randomized clinical trials (RCT) provide a platform to determine whether there is a genuine scientific mechanism associated with the efficacy of a treatment under evaluation (Feinstein 1985). Such trials employ strict design criteria, often controlling for confounding factors that may impact how the study interventions perform. To this end, RCTs rarely reflect actual clinical conditions, and therefore, results fall short of determining how efficacious and safe a treatment may be when used in everyday clinical practice.

In contrast to using placebo, a pragmatic model argues that using currently accepted interventions is the most informative research comparison in a RCT (Schartz 1967). For example, a conventional therapy may be compared with conventional therapy augmented by CAM intervention or a CAM intervention alone. The pragmatic position acknowledges that both intervention-specific and nonspecific factors influence outcomes and therefore, it is difficult to make a clear separation between the effects of an active intervention and that of placebo. Such interactions between intervention-specific and nonspecific effects may even produce unique clinical outcomes (Uhlenhuth 1966). Two analyses have determined that active interventions and placebo do demonstrate dramatically different and unpredictable effects in masked RCTs (Rochon 1999; Kaptchuk 2001a).

Thus, the studies using the pragmatic, interactive approach and those using the placebo-controlled approach are designed to answer distinct questions:

- Pragmatic approach:
  - Questions the assumption that intervention-specific and nonspecific effects are stable, separable, linear, and relatively constant during the duration of a RCT (Spilker 1991).

- Changes the question from whether an intervention is or is not better than placebo towards a question that investigates the magnitude of an effect (Gotzsche 1994).
- Provides less scientifically useful information, but potentially more clinically relevant information.
- Has intrinsic design bias that tends towards the conclusion of no difference between groups unless the sample size is sufficiently large (Jones 1996). In addition, multiple RCTs for standard active interventions have not invariably demonstrated effects greater than placebo, and one cannot be sure that future RCTs will yield superior results unless a placebo arm is included. (Temple 2000).

Designing Credible Placebo in CAM Research: What are the Challenges?

There are many special challenges when designing credible placebo in CAM research. For example, many CAM interventions in the field of physical therapy involve therapist procedures (i.e., chiropractic, massage, etc.) which make concealment and dummy control difficult. Similarly, many of the substances given in CAM are more elaborate than simple chemical structures of drugs often evaluated in traditional clinical trials and pose challenges with regard to concealment. The following examples present strategies designed to optimize concealment used in 4 disparate genre of CAM modalities.

- Oral ingestion of substances
  - Botanicals
    - Bensoussan and colleagues conducted a randomized, double blind, placebo-controlled trial in which patients received either an individualized Chinese herbal formulation, a standard Chinese herbal formulation, or placebo for the treatment of irritable bowel syndrome (Bensoussan 1998). All patients received the same number of pills and were evaluated regularly by a traditional Chinese herbalist and by a gastroenterologist. The placebo, which tasted, smelled, and looked similar to the herbal formulas, comprised a complex formulation of inert substances: 78.2% calcium hydrogen phosphate, 19.6% soy fiber, 0.3% cosmetic brown; 0.5% cosmetic yellow; 0.01% edicol blue; 0.09% identical licorice dry flavor, and 0.03% bitter flavor.
  - Homeopathy
    - The use of placebo in homeopathic research is probably better developed than for any other CAM modality. This is underscored by a meta-analysis conducted by Linde and colleagues that identified 133 placebo-controlled trials involving homeopathy therapy (Linde 1997). One of the challenges associated with homeopathic preparations is related to its peculiar diagnostic system and the multitude of intervention approaches for patients with identical biomedical conditions or diseases. One example of how this challenge was addressed can be found in a report by Fisher and colleagues who evaluated homeopathic treatment in patients with fibrositis (Fisher 1989).
- Use of minimally invasive procedures
  - Acupuncture
    - Placebo acupuncture has evolved from the insertion of a needle at a non-indicated acupuncture point for a condition to the development of sham techniques that involve the use of specially designed telescoping needles with blunt points. This latter

advancement produces a distinct sensation at the point of contact to accompany the appearance of needle insertion (Park 2001; Streitberg 1998).

- Manipulation of "subtle" energies or energies not detected by normative sciences
  - Reiki
    - Reiki practitioners claim that Reiki reduces a variety of physical problems and improves psychospiritual well-being by using 13 hand positions to theoretically reconnect the flow of the patient's universal life energy that has been disrupted by illness. Mansour and colleagues were able to develop and successfully test a sham procedure that including matching not only the hand movements, but also the physical characteristics and demeanor of a trained Reiki master (Mansour 1999). Results of actual efficacy trials using this genre of placebo are not yet available.
- Physical manipulation of the patient's anatomy
  - Chiropractic
    - Chiropractic manipulation is the most popular professional CAM modality in the United States. Some of the sample sham procedures include adjustments using minimal force; an adjustable treatment bench for non-manipulative positional changes; hands-on techniques with no manipulation; and using an adjusting instrument (i.e., activator) with flexible force settings.

Ethical Considerations

The ethics of conducting placebo-controlled trials has been challenged by the recent potential revisions of the Declaration of Helsinki which calls for testing of any new treatments to be done against the best current method where that exists, and not against a placebo (Tollman 2001). Advocates of placebo-controlled trials argue that more benefit than harm is conferred in such trials and that placebo-controlled trials are the foundation for scientifically valid research, and this, therefore, constitutes a fundamental ethical protection (Emanuel 2000). Some argue that the values of patients and community need to be considered when attempting to weigh benefits and risk. In addition, the goal of some trials may not be to show equivalence or superiority over existing therapy, but rather to demonstrate that an intervention provides a positive effect. This may be relevant when issues of toxicities, cost, psychosocial beliefs, or access to care are impacting treatment decisions. Thus, clinicians and researchers need to acknowledge that the issues of placebo controls are complex and be aware of the significance of the options.

COMPLEMENTARY AND ALTERNATIVE MEDICINE (CAM) CLINICAL TRIALS FOR SYMPTOMS: DEVELOPMENT OF APPROPRIATE INTERVENTIONS

Designing clinical trials with CAM interventions can be done without identification of the underlying mechanisms of action for each intervention, as long as a clear and clinically relevant endpoint is used. Nahin and colleagues suggested 3 approaches in which randomized CAM clinical trials can be conducted (Nahin 2001):

Lonnie Zeltzer, MD, David Geffen School of Medicine at the UCLA, author

Christine Miaskowski, RN, PhD, University of California San Francisco, discussant

- Study the whole intact traditional system to treat a specific disease.
  - Example: A current study compares traditional Chinese medicine (TCM), naturopathic medicine, and conventional care in the treatment of patients with temporomandibular disease.
- Study a specific modality adapted from a traditional system for treating a specific disease.
  - Example: Ongoing trial of the use of acupuncture for patients with depression in which individualized acupuncture treatments based on TCM diagnosis and points for depression are compared to other acupuncture points and a wait list control group.
- Test a single standardized intervention.
  - Example: Use of an herbal medicine to treat a conventionally diagnosed disease.

In conducting CAM trials, consideration should be given as to whether a standardized intervention or an individualized approach is to be tested and whether these approaches are being evaluated for a single component (e.g., disease) or a system (i.e., whole person). The advantages and disadvantages of the individualized approaches in each setting are outlined in Table 1.
The optimal design of CAM interventions in clinical trials would be one using an integrated approach comprising a standardized intervention with some room for individualization within the general CAM treatment. This approach would preserve both external and internal validity of outcomes. Examples of this approach include:

- Standard yoga sequences of poses for neck pain that can be individualized within the general treatment.
- Reiki energy therapy where standardized techniques such as magnetic unruffling or specific hand positions would be used, but the application would be individualized based on the patient's needs.

Studies that integrate individualization within a standardized framework, with details of the individual treatments recorded, will offer the best of both approaches. To this end, the future of CAM research will benefit from shared experience of CAM clinicians, biomedical researchers, and social scientists. This cross-fertilization will enhance the scientific rigor of research while preserving the philosophic basis of CAM and its applicability to the clinical setting.

Table 1. Potential Methodologic Advantages and Disadvantages of Individualization of CAM Intervention

	Advantages	*Disadvantages*
Study of CAM component	• Can most accurately reflect how a CAM modality is practiced in the community • CAM may produce greater benefit because it is geared toward the individual patient and embedded in the intent of the CAM practice • If effect sizes are large, then sample sizes can be smaller • Can maximize effect size and then conduct a next-step study by comparing the individualized approach to a standard approach • Optimized by including the CAM-based diagnosis and examining outcomes in relation to both Western and CAM diagnosis	• Design of control group is difficult • Design of a double-blind study may be less feasible compared to a standardized approach • May be more difficult to start with a controlled, standardized approach, and if effective, conduct a next-step comparison with an individualized approach to treatment • Does not easily allow for internal validity and generalization of the intervention as compared with a standardized approach • Harder to minimize physician differences compared with standardized approaches • Group interventions are less amenable compared with a standardized approach • Underlying mechanisms are less easily assessed compared with a standardized approach
Study of CAM systems	• Can most accurately reflect how a CAM intervention is practiced in the community, thereby addressing external validity • May maximize main effects in a shorter amount of time than use of components of the CAM intervention • If effect sizes are large, then sample sizes can be relatively smaller • Can be compared to other CAM systems or standard treatment, with side effect profile and cost effectiveness considered in outcomes • More likely to get cooperation from CAM clinical community because they are often frustrated when research evaluates CAM modalities out of context	• Difficult to design appropriate controls • Need a series of dismantling studies to determine whether the CAM intervention as a whole is needed or whether there are critical components and which of those components is/are responsible for the outcome • May be difficult for CAM clinicians to agree upon all the different components: more decisions go into the description and the adherence to the treatment

ASSESSMENT OF CANCER-RELATED SYMPTOMS: RELEVANCE FOR CAM RESEARCH METHODOLOGY

Charles S. Cleeland, PhD, University of Texas MD Anderson Cancer Center, author

Patricia A. McGrath, PhD, The University of Western Ontario, discussant

Cancer-related symptoms can significantly impair the daily function and quality of life of patients as well as negatively influence disease-related outcomes. Despite the tremendous distress symptoms can cause, symptom assessment is rarely a part of routine cancer care. It is no surprise, therefore, that the lack of systematic assessment and treatment of symptoms makes the relief of cancer-related symptoms one of the most frequent reasons why patients seek out CAM.

Thus, the widespread use of CAM and the promise of many CAM-based interventions for cancer-related symptom relief underlines the necessity for well-designed clinical trials to evaluate and determine the clinical utility of various modalities. A critical factor of such trials, however, is the use of reliable, valid, and practical measures of symptom severity and impact.

General Guidelines for Symptom Assessment

The use of standardized assessment instruments provides patients with a non-threatening format to report symptoms. Simple measurement scales can greatly improve symptom assessment, direct treatment choices, and assess the effectiveness of treatment. For example, most evidence-based guidelines and clinical pathways for the management of cancer pain assume that pain severity scores will be used to assess pain and monitor treatment progress (Max 2002).

An ideal symptom assessment tool for cancer patients should include disease and treatment-related symptoms that are both common and distressing for patients. Symptom severity can be rated using Visual Analog Scales, Verbal Rating Scales, and Numeric Rating Scales. These scales have been extensively used in pain research and provide nearly equivalent data.

Although single-symptom scales can be useful in documenting symptom severity and guiding treatment, most cancer patients with advanced disease experience multiple symptoms, either from cancer or treatment. In addition, many potential treatments, including some CAM treatments, are thought to have a general positive effect for the patient, but can only be assessed using instruments that cover multiple symptom domains. Some of the more common multi-symptom scales for assessing cancer-related symptoms are listed below:

- Symptom Distress Scale (SDS)
  - Originally developed by McCorkle in 1978, a number of modifications have been made to customize the SDS to individual cancers.

- Memorial Symptom Assessment Scale (MSAS)
  - The MSAS evaluates 32 physical and psychological symptoms and has been used to provide a comprehensive assessment of the prevalence and characteristics of a wide spectrum of symptoms.
- Rotterdam Symptom Checklist (RSC)
  - The RSC is a 30-item scale designed to measure cancer symptoms in patients who participate in clinical trials. This scale does not include pain as a general symptom.
- Edmonton Symptom Assessment System (ESAS)
  - Designed for cancer patients in palliative care, the ESAS comprises 9 Visual Analog Scales that measure pain, activity, nausea, depression, anxiety, drowsiness, appetite, sensation of well-being, and shortness of breath.
- M.D. Anderson Symptom Inventory (MDASI)
  - The core MDSAI consists of 13 symptoms (pain, fatigue, nausea, sleep disturbance, emotional distress, shortness of breath, lack of appetite, drowsiness, dry mouth, sadness, vomiting, difficulty remembering, and numbness or tingling), but may also include modules of additional symptoms for those patients who are at risk for symptoms not highly prevalent in oncology patients in general.

Assessment of individual cancer-related symptoms can be useful when additional information about a specific symptom is desired, especially in cases where symptom treatment is difficult and/or very distressing for the patient. There are a number of assessment instruments for cancer-related pain, fatigue, and depression. When a single symptom does not have a validated assessment instrument, clinicians can use the individual symptom item from a multiple-symptom scale.

In summary, many CAM-based treatments have symptom relief as a target for patients with cancer. General recommendations for designing clinical trials to assessing symptoms in cancer patients include:

- Generate a reasonable hypothesis about what might happen.
- Consider what a reader might want to learn from that trial.
- Consider patient burden, favoring assessment instruments that are short and easy to understand.
- Keep the trial as short as possible, but allow enough time to answer questions raised by the study hypothesis.
- Consider using scales that assess multiple symptoms.

THEORY, HYPOTHESIS DEVELOPMENT, AND RANDOMIZED CLINICAL TRIALS IN COMPLEMENTARY AND ALTERNATIVE MEDICINE: A BEGINNING AT SEPARATING SCIENCE FROM POLITICAL SCIENCE

Gary R. Morrow, PhD, University of Rochester Cancer Center, author

Richard J. Gralla, MD, Columbia-Presbyterian Medical Center, discussant

The testing of new therapeutic agents progresses through 3 phases of adequate and well-controlled clinical trials that are designed to determine safety and efficacy of a potential new therapy. Under the auspices of the United States Food and Drug Administration (FDA), the progression among the phases are:

- Phase I
  - Designed to gather preliminary data on a new drug's safety using a small number of healthy or ill subjects. Estimated time to complete this phase is 1 year.
- Phase II
  - Designed to determine the most effective dose level for efficacy as well as continue evaluating the drug's safety. Estimated time to complete this phase is 2 years.
- Phase III
  - Designed to evaluate the safety and efficacy of the new drug in a large number of patients with a particular disease or condition. Patients are randomized to either the new drug or the current standard of treatment. Placebo is used if no standard treatment exits. Estimated time to complete this phase is 3 years.

Once a New Drug Application is submitted, the FDA usually completes its review and makes a recommendation within 6 months to 2.5 years. In addition, the FDA has increased its surveillance of a drug's safety and efficacy by strongly encouraging the manufacturer to conduct post-marketing (Lasagna 1989). In general, oncology research has closely followed the requirements specified by the FDA and the review time for oncology drugs has improved over recent years.

Because Phase I/II studies are generally focused more on safety and dosing issues, the usefulness of such studies of CAM modalities might need to be examined. This is due to the generally less toxic attributes of CAM. However, dosing issues with CAM modalities can present a considerable challenge, especially with regard to herbal preparations. The ideal CAM research design, as in conventional medicine, remains the double-blind, randomized, placebo-controlled trial. Because this model does not fit all CAM modalities, an increased emphasis should be placed on the replication of research results by several independent researchers. To this end, CAM experiments must be designed and described in detail sufficient to make replication possible. As standards and common methodologies evolve, it is useful to remember that virtually all currently accepted medical interventions were seen as unproven, complementary, and alternative at one time.

DEVELOPMENT OF BOTANICAL "NEW DRUGS": FDA REGULATIONS

Shaw T. Chen, MD, PhD, Food and Drug Administration, presentation

Botanicals can be marketed in the United States as either a food, a dietary supplement, an over-the-counter product, or as a drug. Issues specific to marketing a botanical drug under a New Drug Application were addressed. Discussion was drawn from the FDA's draft Guidance for Industry: Botanical Drug Products, published in August 2000 and available at www.fda.gov/cder/guidance/1221dft.pdf. Final publication is expected second quarter 2003.

The current FDA policy for botanicals requires that the safety and efficacy of each component be characterized and the optimal ratio/doses of each component investigated. This requirement has changed in the recent draft guidance paper. Identification of the active constituents is not essential and the Chemistry, Manufacturing, and Control (CMC) requirements will be extended to controls of raw materials. Another important feature in the new guidance paper is that pre-clinical evaluation can occur concurrently with (or later than) clinical studies.

The timing, sequence, and/or extent of preclinical testing will be adjusted based on what is already known about a particular botanical. The requirement for an Investigational New Drug (IND)/New Drug Application (NDA) depends on several factors including the past history of the botanical, scales of the trials previously conducted, the degree of modification from past formulations, and the novelty of the therapy. The guidance is divided into 2 major categories: preliminary studies (i.e., phase I/II) and expanded studies (i.e., phase III).

Some practical advice about how to develop botanicals as new drugs from the regulatory perspective includes:

- The NDA should include:
  - Detailed CMC characterization
  - Assurance from preclinical safety data
  - Adequate and well controlled clinical studies
- The IND should include:
  - Evidence of prior use
    - Documentation of US marketing
    - Extensive review of past experiences
    - Balanced summary of current use
    - All types of old data will be considered
  - CMC information
    - Control of raw materials and process
    - Certain tests along with clinical development
    - Use of a single source, process, dosage form, and batch
    - Retain materials for future tests
  - Pharmacology/toxicology data
  - Clinical development plan

The clinical considerations associated with initiating a clinical trial with a botanical are the same as for trials with purified chemical drugs. First is the use of an established diagnosis and clinical endpoint. The roles of an alternate medical theory may also be incorporated into the study design. The ease with which both patient and physician can understand the specifics of the trial is another consideration. Applicants are strongly encouraged to consult with the FDA for study designs.

The need for an IND for a dietary supplement already lawfully marketed in the United States depends on the intended use. If the new botanical in question is to be used to treat, diagnose, or prevent a disease, that is considered a disease use and an IND is required. An IND is not required to conduct a pharmacodynamic study in patients if the data are to be used for scientific reasons. For studies that require an IND, the FDA will review protocol objectives, endpoints, informed consent documents, and patient perception issues. Advantages of going through the IND process include increasing the likelihood that the data are useful by ensuring consistency in product quality, demonstrating that the product already meets pharmacology/toxicology standards, thus resulting in clinical studies that are well-designed and properly implemented.

ETHICAL CHALLENGES IN COMPLEMENTARY AND ALTERNATIVE MEDICINE CANCER SYMPTOM RESEARCH

Jacquelyn Slomka, PhD, The University of Texas Health Sciences Center at Houston, author

In general, the ethical issues facing CAM research are similar to those encountered in any clinical research endeavor. Guidance in ethical conduct for clinical trials have been shaped by many initiatives, including the:

- Nuremberg Code (available at: http://ohsr.od.nih.gov/nuremberg.php3. Accessed 1/6/03)
- World Medical Association's Declaration of Helsinki (available at: http://ohsr.od.nih.gov/helsinki.php3. Accessed 1/6/03)
- Council for International Organizations of Medical Sciences (available at: http://www.cioms.ch/frame_guidelines_nov_2002.htm. Accessed 1/6/03)
- Belmont Report (available at: http://ohsr.od.nih.gov/mpa/belmont.php3. Accessed 1/6/03)
- National Bioethics Advisory Commission (available at: http://www.georgetown.edu/research/nrcbl/nbac. Accessed 1/6/03).

However, CAM research in the setting of cancer symptom management at the end of life presents special ethical concerns that are specific to the nature of CAM therapies. These concerns include the use of CAM interventions in potentially vulnerable patient populations (e.g., children and terminally ill cancer patients), the perceived view among many patients and healthcare professionals of CAM being unorthodox, and the potential financial benefit CAM therapy may have over orthodox treatments.

Methodology: CAM and Ethics

The issue of how best to evaluate CAM therapy extends to the ethical realm. Physicians are obliged to offer patients therapeutic options that are safe and effective, often considering therapies that have been evaluated using conventional methods (i.e., RCT) (Emanuel 2000). It is well recognized that methodologic issues confound interpretation of results from CAM trials, and place in doubt the efficacy and safety of CAM interventions. As long as conventional methods of evaluation remain the standard for assessing unconventional treatment, ethical concerns of offering CAM interventions will remain.

Clinical Equipoise and CAM

Clinical equipoise, the state of the medical community's lack of consensus about the comparative merits of a therapy before testing it, is an ethical requirement of clinical research. Whether or not the individual physician is able to recommend or discourage the patient's participation in a clinical trial of a controversial treatment is irrelevant because the state of equipoise derives from the medical community and not from an individual physician or investigator (Freedman 1987). Thus, one ethical problem for research on CAM is that true clinical equipoise will be very difficult if the medical community's opinion remains weighted against CAM. The current legal climate of health care delivery may be one factor contributing to this negativity (Cohen 2002).

However, such equipoise is not impossible and studies do proceed, as in the case of folic acid use to prevent neural tube defects or zinc to treat colds.

Ethics and Placebo

The research community remains divided about the ethics of placebo use in clinical trials when effective treatment exists, though some have attempted to bridge their differences by applying ethics criteria on a case-by-case basis (Emanuel 2002). The general controversy over placebo use is relevant to CAM research, especially in terminally ill cancer patients, where the physical and psychological effects of withholding known effective treatment may be more severe than in healthier patients (Markman 1994). This underscores the importance of providing true informed consent among clinical trial participants of the potential burdens caused by withholding effective treatment. Ultimately, the issue of harm and the use of placebo in CAM trials may indeed have to be considered on a case-to-case basis.

Use of Vulnerable Persons as Research Participants

Children and the terminally ill cancer patient are two of the vulnerable populations that may be impacted by CAM research. Because little is known about CAM research in children, investigators may need to conduct animal studies, phase I trials, or related adult or adolescent studies prior to testing in younger children (Personal communication, Dr. Lonnie Zeltzer, 2002). For terminally ill cancer subjects, symptoms such as fatigue and change in mental status may present barriers to clinical trial participation. It is, therefore, incumbent upon CAM investigators to demonstrate the value of CAM research as a legitimate adjunct to standard medical research before considering trials in children (Personal communication, Dr. Patricia McGrath, 2002) and terminally ill patients.

In publishing research results, ideally investigators should consider describing how vulnerable patients were recruited and enrolled in the study to clarify the consent process and ensure its validity. Additional consent issues to consider for vulnerable populations include:

- Designation of a proxy decision maker in cases when full consent of the patient is not possible.
- Advanced directives that address issue of continued participation when the subject is unable to consent or withdraw (Dresser 1996).
- Community-based consent in which standards for participation are developed through community consensus (Kraybill 1999).

Conflict Loyalties

Potential conflicts exist among physicians who assume dual roles of caregiver and investigator. Additional questions of conflict arise when corporate support of medical research is involved. Thus, physicians' and researchers' loyalties are shared with industry, agencies, and corporations that fund research. Currently, physicians and researchers are obligated to disclose possible conflicts of interest to maintain the trust of patients and peers in the research process.

Future Areas of Research

Six areas of further research are suggested:

- The role of researcher conflict of interest and its effect on the physician-investigator-patient-subject relationship.
- The effect of professionals' attitudes and perceptions about CAM on the research process.
- The requirement for clinical equipoise in research.
- Protection of vulnerable subjects in CAM/palliative care research.
- Special issues in the design of CAM trials.
- "Controlling for culture" in clinical trials.

STATISTICAL ISSUES IN THE STUDY OF CAM INTERVENTIONS

Jeff A. Sloan, PhD, Mayo Clinic, author

Andrew Vickers, DPhil, Memorial Sloan-Kettering Cancer Center, discussant

Existing statistical methods and a sound scientific paradigm are the only prerequisites necessary to design sound clinical research on the possible benefits and safety of CAM interventions. The hurdles to the development of appropriate guidelines from existing research for clinical use of CAM interventions lie in the classical problems of imprecision of definitions of CAM interventions, the use of subjective measures, and the reality of multiple outcomes.

The primary challenge encountered in clinical research of CAM interventions is inability to delineate exactly what the therapy is, what it is expected to do, and its mechanism of action (Table 2).

Fortunately, most of these problems are easily addressed. It is essential that the collective body of evidence gained from multiple statistical approaches should be considered so that complete and sufficient evidence can be collected. A proposed feasible approach for conducting CAM studies and approaching potential hurdles is outlined in Table 3.

The difficulties encountered in CAM research appear to be in two areas. First is the translation from research question to study design and measurement tools used. The other is the consequent implications for the appropriate statistical analysis and the communication of results and their interpretability to clinical practice. The proposed model provides guidance on conducting scientifically rigorous clinical studies for evaluating any therapy, including CAM interventions.

Table 2. Typical Problems Encountered During CAM Research

At the Study Design Phase

- The collection and analysis of data often precedes the appropriate definition of primary and secondary endpoints.
  - Mixed results are difficult to interpret because there is no clarity regarding the primary endpoint.
- Measurement tools and items do not focus on the identified outcomes because of poor choice of measurement tools.
- The psychometric properties of new measures or items from existing tools are not established before use.
- The comparability of study results is often hindered due to the lack of consistency between measurement scales and approaches of the various measurement tools used.
  - Some form of data transformation is needed to make comparisons.
- Inappropriate or missing power analysis makes it unclear if the conclusion of the study is a function of mere sample size or a true indication of relationships among dependent and independent factors.
- The magnitude of the observed effects can not be interpreted in terms of clinical significance because the appropriate sample size was not determined in advance.
  - Clinical practice recommendations can not be derived from the study results.
- Study is not adequately stratified to account for potentially confounding variables.
  - True impact of the CAM intervention on the endpoint is difficult to determine.

At the Statistical Analysis Stage

- Significant patient attrition or missing data can compromise application of standard statistical methods.
- Imbalances between the size of the treatment arms are often not taken into account in the statistical analysis.
- Threshold values that are likely to change the study conclusions are often not explored using sensitivity analysis. This is particularly important for new or merging therapies.
- Not enough data are available to sustain the number of variables included in the statistical model.

At the Stage of Presentation and Interpretation of Results

- Reported results often are focused on averages and fail to include the distribution and range of values. These latter data are as important as mean values, especially for newer therapies like CAM.
- Typically, the statistical significance of results are emphasized with no consideration given to the clinical significance of study findings.

Table 3. Proposed Model for Conducting CAM Clinical Studies

At the Study Design Phase

- Appropriate *a priori* linkage must be established between the:
  - Research hypothesis,
  - Chosen endpoints,
  - Measurement tools,
  - Study design,
  - Statistical analysis.
- A single primary endpoint and appropriate secondary endpoints must be identified in advance.
- Study design and planned statistical analysis must be focused on the primary endpoint.
  - Proper specification and sample size must be defined for the primary endpoint.
  - Appropriate statistical analyses must be identified for each of the secondary endpoints.
  - Multiple endpoints must be accounted for by appropriate statistical approach.
- Potential confounding variables must be identified and accounted for in terms of appropriate stratification.

At the Statistical Analysis Stage

- Appropriate statistical tests must be identified for primary and secondary endpoints.
  - Analysis of the primary endpoint must be kept separate from supporting analyses.
- Missing data must be examined for patterns of attrition, specifically, whether data are missing at random or in a systematic fashion. Adjustments for missing data must be made accordingly.
- The assumptions underlying chosen statistical procedures must be examined for validity.
- A comprehensive sensitivity analysis using different statistical procedures, with their underlying assumptions must be conducted to establish the robustness of findings.
- Intent-to-treat analysis should be the statistical approach.

Separate analyses on appropriately pre-specified subpopulations must be conducted.

At the Stage of Presentation and Interpretation of Results

- Descriptive as well as distribution statistics must be presented for primary and secondary variables.
- Study results must include results pertaining to the "average" and "typical" patient.
- Graphical displays of results for individual patients must be presented, if possible, to demonstrate the range of effects over study sample.
- Results from basic statistical analysis must precede the presentation of results from more complex modeling.
- Discrepancies in findings between basic and complex modeling must be discussed and explained.
- Sensitivity and threshold analyses must be presented and discussed.
- The statistical significance of results must include interpretation of *p*-values.
  - The clinical significance of results must be presented and discussed.

STRATEGIES FOR GRANT PREPARATION IN CANCER CAM SYMPTOM RESEARCH

Strategies for Applicants in Cancer CAM Symptom Research

Placebo/Shams/Control Groups:

Use placebo to demonstrate whether a therapeutic intervention has effect.
Use active comparison to demonstrate how strong an effect an intervention may have.
Create placebos and shams as similar as possible to intervention.
Defend strategy of including or not including comparison groups.

***Individualized or Standardized Approach to CAM Interventions*:**

Discuss advantages and disadvantages of each approach.
Provide compelling rationale for choice.
Consider integrating individualized approach within standardized format.

Measurement Issues:

Include hypotheses/rationale about why intervention would affect these symptoms.
Use standardized tools that have demonstrated validity and reliability.
Use tools that measures most common and most distressing symptoms.
Consider tools that measure multiple symptoms.
Consider and address patient burden.

***Selecting Phase*:**

Defend proposing Phase III without Phase I or Phase II data—does "thousands of years use"suffice?
Address dosing issues—if don't know dosage information, get preliminary data.
Give enough detail for replication.

Investigational New Drug (IND) Issues:

May require IND even if available over the counter—depends upon use.
For NIH proposals, INDs may not be required—contact FDA and NIH program staff to inquire.
Phase I/II studies may not require preclinical data: Phase III may require more toxicity data.
INDs encouraged, as the process can improve study design and increase likelihood of usable data.

Ethics:

Demonstrate value of CAM research as a legitimate adjunct to conventional medical research.
Disclosure of conflict of interest to patients is essential.
Describe how vulnerable patients are recruited and enrolled to clarify and ensure informed consent.

Statistical Issues:

Define primary and secondary endpoints.
Choose measurement tools that focus on those endpoints.
Include appropriate power analysis.
Use stratification to account for confounds.
Detail how to address patient attrition and/or missing data.
Discuss both statistical significance and clinical significance.

REFERENCES

REFERENCES

Bensoussan A, Talley NJ, Hing M, Menzies R, Guo A, Ngu M. Treatment of irritable bowel syndrome with Chinese herbal medicine: a randomized controlled trial. *JAMA*. 1998;280:1585-1589.

Cohen MH, Eisenberg DM. Potential physician malpractice liability associated with complementary and integrative medical therapies. *Ann Intern Med*. 2002;136:596-603.

Dresser R. Mentally disabled research subjects. The enduring policy issues. *JAMA*. 1996;276:67-72.

Emanuel EJ, Miller FG. The ethics of placebo-controlled trials—a middle ground. *N Engl J Med*. 2001;345:915-919.

Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA*. 2000;283:2701-2711.

Feinstein, AR. *Clinical Epidemiology*. Philadelphia: WB Saunders, 1985.

Fisher P, Greenwood A, Huskisson EC, Turner P, Belon P. Effect of homeopathic treatment on fibrositis (primary fibromyalgia). *Br Med J*. 1989 Aug 5;299(6695):365-366.

Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*. 1987;31:141-145.

Gotzsche PC. Is there logic in the placebo? *Lancet*. 1994;344:925-926.

Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *Br Med J*. 1996;313:36-39.

Kaptchuk TJ, Eisenberg DM. Varieties of healing. 1: medical pluralism in the United States. *Ann Intern Med*. 2001;135:189-195.

Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med*. 1998;72:389-433.

Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *J Clin Epidemiol*. 2001a;54:541-549.

Kraybill EN, Bauer BS. Can community consultation substitute for informed consent in emergency medicine research? In: King NMP, Henderson GE, Stein J, eds. *Beyond Regulations. Ethics in Human Subjects Research.* Chapel Hill, NC: The University of North Carolina Press, 1999:191-198.

Lasagna L. Congress, the FDA, and new drug development: before and after 1962. *Perspect Biol Med*. 1989;32:322-343.

Linde K, Clausius N, Ramirez G, et al. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet*. 1997;350:834-843.

Mansour AA, Beuche M, Laing G, Leis A, Nurse J. A study to test the effectiveness of placebo Reiki standardization procedures developed for a planned Reiki efficacy study. *J Altern Complement Med*. 1999;5:153-164.

Markman M. When regulatory requirements conflict with ethical study design: the case of oral ondansetron. *Cancer Invest*. 1994;12:654-656.

Max MB, Payne R, Edwards WT, Inturrisi CE, Sunshine A. *Principles of Analgesic Use in the Treatment of Acute Pain and Cancer Pain.* 4th ed. Glenview, IL: American Pain Society; 2002.

Nahin RL, Straus SE. Research into complementary and alternative medicine: problems and potential. *Br Med J*. 2001;322:161-164.

Park J, White AR, Ernst E. New sham method in auricular acupuncture. *Arch Intern Med*. 2001;161:894.

Rochon PA, Binns MA, Litner JA, et al. Are randomized control trial outcomes influenced by the inclusion of a placebo group?: a systematic review of nonsteroidal antiinflammatory drug trials for arthritis treatment. *J Clin Epidemiol*. 1999;52:113-122.

Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20:637-648.

Spilker B. *Guide to Clinical Trials*. New York, NY: Raven Press, 1991.

Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet*. 1998;352:364-365.

Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med*. 2000;133:455-463.

Tollman SM, Bastian H, Doll R, Hirsch LJ, Guess HA. What are the effects of the fifth revision of the Declaration of Helsinki? *Br Med J.* 2001;323:1417-1423.

Uhlenhuth EH, Rickels K, Fisher S, Park LC, Lipman RS, Mock J. Drug, doctor's verbal attitude and clinic setting in the symptomatic response to pharmacotherapy. *Psychopharmacologia*. 1966;9:392-418.

ORIGINAL PAPERS BY PANELISTS

The Placebo and Complementary and Alternative Medicine: Historical, Scientific, and Methodological Considerations 25

Complementary and Alternative Medicine (CAM) Clinical Trials for Symptoms: Development of Appropriate Interventions 39

Assessment of Cancer-Related Symptoms: Relevance for CAM Research Methodology 57

Theory, Hypothesis Development, and Randomized Clinical Trials in Complementary and Alternative Medicine; A Beginning at Separating Science from Political Science 79

Ethical Challenges in Complementary and Alternative Medicine Cancer Symptom Research 91

Statistical Issues in the Study of CAM Interventions 103

THE PLACEBO AND COMPLEMENTARY AND ALTERNATIVE MEDICINE: HISTORICAL, SCIENTIFIC, AND METHODOLOGICAL CONSIDERATIONS

Ted J. Kaptchuk, OMD*
Associate Professor of Medicine
Osher Institute
Harvard Medical School

Brian M. Berman, MD
Director, Complementary Medicine Program
University of Maryland School of Medicine

R. Barker Bausell, PhD
Director of Research, Complementary Medicine Program
University of Maryland School of Medicine

**Corresponding author:*
Ted J. Kaptchuk, OMD
Osher Institute
Harvard Medical School
Boston, Massachusetts 02215
617-384-8550
ted_kaptchuk@hms.harvard.edu

THE PLACEBO AND COMPLEMENTARY AND ALTERNATIVE MEDICINE: HISTORICAL, SCIENTIFIC AND METHODOLOGICAL CONSIDERATIONS

SUMMARY:

The placebo and the issue of placebo effects have been integral to the debate on alternative therapies since the development of a fissure between mainstream and unconventional medicine. In fact, while placebo controls were only adopted for routine evaluation of mainstream therapies after World War II, placebo controls themselves were first developed to test unconventional therapies over one hundred and fifty years before. While the use of placebo controls in any medical research often poses challenges, their use in complementary and alternative medicine (CAM) research can be especially difficult. Besides providing an historical background, this paper addresses issues of placebo controls in CAM research. Questions discussed include: 1. Are placebo controls necessary for establishing the efficacy of conventional and alternative medicine (CAM)? 2. What are some of the unique challenges in designing credible placebo interventions in CAM? 3. In fact, is there really a "powerful" placebo effect? 4. Is it possible that there is an "enhanced" CAM placebo effect? and 5. What are the ethical issues concerning placebo controls in clinical trials? Simple answers for these questions that command universal consensus do not exist. An awareness of the complexity of the placebo issues in CAM may contribute to greater understanding on the part of clinicians and more rigorous research on the part of researchers.

HISTORICAL BACKGROUND:

One of the most frequently asked questions, concerning CAM therapies is whether they are anything more than placebo effects. In fact, since the time mainstream and alternative medicine bifurcated some two hundred years ago, the placebo has been integral to the rhetorical tug-of-war between the two camps.[1] Frequently, polemical attacks on unconventional medicine included the accusation that alternative medicine was no more than a "dummy pill." For example, in 1852, the predecessor journal of the *New England Journal of Medicine*, claimed that the "miserable chicanery" of then-popular forms of "quackery" were all based on "placeboism."[2] In 1946, at the first scientific conference on placebos, it was claimed that the success of alternative medicine demonstrated very clearly what can be done by placebos."[3] Contemporary debate continues to discern the footprint of placebos in CAM[4,5,6] and many "regard alternative medicine much the way [they] do a placebo [treatment]"[7] composed of "a cadre of therapists delivering nonspecific effects."[8]

In fact, the origin of one current conceptual biomedical model that defines efficacious or "legitimate" therapy as an intervention that is more than an indistinguishable placebo treatment probably arose in the historical conflict between conventional and mainstream medicine. Almost 200 years before routine medical research adopted placebo controls, researchers concerned with alternative medicine (as either neutral scientists, debunkers, or advocates) understood the need for and developed placebo controls. When "preposterous" mechanisms of cure or relief were advocated, mainstream scientists (and then unconventional practitioners) saw the necessity to utilize the "extra precaution" of placebo controls.[9] In the history of medical research, the first

placebo sham device was introduced in the mesmerism/magnetism controversies at the end of the eighteenth and beginning of the nineteenth centuries. The first adoption of placebo pills to mask a verum (active treatment) took place during the homeopathic controversy beginning in the middle of the nineteenth century.[9] These decoy treatments were adopted in response to the accusation that the perception of efficacy on the part of these unconventional therapies had to do with the appearance of an intervention as opposed to any specific content. Researchers felt that the implausible mechanism claims of alternative medicine required controlling for the possibility that any improvement seen in treatment might be due to "imagination," "suggestion," or "natural history."

Before World War II placebo controls (and even concurrent controls) were generally absent in mainstream medical research.[10] Efficacy was defined in terms of clinical outcomes and changes made in patients' baseline condition. After World War II, reacting to the crisis due to the rapid introduction of many new drugs, conventional medicine quickly accepted the placebo methodology that had been pioneered in alternative medicine research. The introduction of this new methodology automatically (and sometimes without awareness) created a shift in the conceptual understanding of what it meant to claim a treatment is efficacious or legitimate. In the new model, one definition of efficacious therapy was operationally taken to be a treatment that produces positive benefits greater than the effects of a placebo intervention in a randomized controlled trial (RCT).[11,12] Clinical outcomes in themselves (changes in patient baseline), became less critical than the fact that the intervention produced a medical outcome significantly greater than a placebo effect in a RCT.[13] Legitimate therapy came to mean the existence, in a predefined cohort of patients, of a demonstrable cause-and-effect relationship between an isolated verum (active) intervention and a predetermined measurable outcome.[14]

ISSUES OF PLACEBO CONTROLS IN CAM RESEARCH

Are Placebo Controls the Only Legitimate Method for Establishing Efficacy?

Placebo controls in RCTs answer the question of whether the medical outcome of the verum treatment group is statistically significantly greater than the placebo effect. This is determined by comparing the verum treatment group with a treatment group that receives a dummy treatment that should mimic the appearance, taste, and feel of the real treatment. Ideally placebo and active treatment should be indistinguishable. Any positive effect in the placebo arm could be ascribed to any of the following considerations: expectation, clinical attention, conditioning, experimental subordination, anxiety reduction, concurrent treatments, nursing care, careful diet and rest, natural history, and regression to the mean.[15] Usually such placebo outcomes are treated as necessary nuisance noise in RCTs but are otherwise considered inconsequential or treated with contempt.[16]

While the placebo-controlled RCT answers the scientific or "fastidious" question of whether there is a "genuine" scientific mechanism,[17] it does not answer necessarily answer the "pragmatic" question of what the intervention will do under normal clinical conditions.[18] Epidemiologists have noted that the magnitude of verum and placebo treatments may be different in RCTs than in actual clinical conditions.[19] Advocates of pragmatic research—who probably have a disproportionate large following in the CAM community—argue that trials are best

carried out in normal or optimal environments with an emphasis on acquiring information necessary for making a clinical decision. This pragmatic model usually argues that using a currently accepted intervention, as opposed to using a placebo, is the most informative research comparison in a RCT. Sometimes this model advocates comparing conventional therapy with conventional therapy augmented by CAM therapy. This position often also claims that various non-specific effects interact among themselves and among any verum-specific effects; and therefore, that it is difficult to make a clear separation between verum and placebo. This approach to placebo, which has been called "interactive," further argues that there is sufficient evidence of synergy between verum-specific and non-specific effects which generate unique clinical outcomes.[20] Some research evidence seems to indicate that active interventions and placebo interventions in fact do have dramatically different and unpredictable outcomes in masked RCTs as compared to RCTs that use active controls.[21,22] The pragmatic, interactive, approach implicitly questions central facets of the fastidious approach including the assumption that the verum-specific and non-specific effects are stable, separable, linear, and relatively constant during the duration of an RCT.[23] Furthermore, the pragmatic approach "switches [the question] from whether or not an intervention is a placebo, towards the magnitude of the effect."[24] A pragmatic perspective provides less scientifically useful information but potentially provides more clinically relevant information. The growth of the outcomes research approach is consistent with this approach. One problem of active-control equivalence or non-inferiority trials is that they have an intrinsic design bias which tends toward the conclusion of no difference unless the sample size is sufficiently large.[25] Also, non-inferiority trials are actually a form of historical control and one can never be sure that the standard therapy will be greater than a placebo in the subsequent comparative trial. (Multiple RCTs for many standard therapies do not invariably demonstrate effects greater than placebo and one can never be sure that the active-control is more than a placebo in an ongoing trial unless a placebo control is also included.[26] See discussion below.) At this time, the medical community, as represented by funding agencies such as the NIH and peer-reviewed journals seem to accept the need and reasonableness of both a fastidious and pragmatic research approach. Researchers and clinicians who depend on research need to be aware of these two approaches answering distinct questions. Continuing the earlier historical debates between mainstream and alternative therapies mentioned earlier, many scientific researchers feel that because the mechanism rationale for CAM therapies is not within normative scientific bounds, the fastidious model of RCT should have priority. Also, many researchers would also argue that placebo controls are critical to insure blind assessment and control for selection and detection bias which undoubtedly could be critical in a controversial arena such as CAM. CAM advocates and clinicians often feel that the pragmatic question is more relevant and critical for patient care.

WHAT ARE THE UNIQUE CHALLENGES OF DESIGNING CREDIBLE PLACEBO IN CAM RESEARCH?

While the techniques for employing placebos in drug trials are relatively simple, special challenges exist for CAM research. Like surgery and physical therapies, many CAM interventions involve therapist procedures (such as chiropractic or massage) which make concealment and dummy control especially difficult. Like the procedures attending social, psychological, and behavioral investigations it is often not immediately apparent how credible controls can be constructed that produce *only* nonspecific effects. Also, unlike the simple

crystalline structures of pharmaceuticals, many of the substances given orally in CAM are more elaborate and pose unique problems in terms of concealment. Nonetheless, as mentioned earlier, given that many attribute the treatment effects claimed by CAM therapies as due *solely* to placebo effects, investigators evaluating each individual therapy must wrestle with techniques for designing control groups that can indeed be assumed to elicit *only* a placebo effect and still maintain masking. In illustration, we will present examples from placebos employed in research evaluating the efficacy of four disparate genre of CAM modalities involving: (a) the ingestion of substances orally (e.g., herbal therapy, homeopathy), (b) the use of a minimally invasive procedure (e.g., acupuncture), (c) the manipulation of "subtle" or energies not detected by normative sciences (e.g., Reiki, therapeutic touch, spiritual healing), and (d) the physical manipulation of the patient's anatomy (e.g., chiropractic, massage).

Botanicals

Botanicals are a serious challenge for developing credible and effective dummy controls. This becomes especially apparent in traditional Chinese Medicine. The following example taken from a study published in the 1998 *Journal of the American Medical Association (JAMA)* coordinated theme issue on CAM can be seen as a model of sophistication and representative of the problem of developing placebo controls in CAM botanical trials. A total of 116 patients with irritable bowel syndrome were randomly allocated to one of three arms: a placebo treatment, a standardized Chinese herbal formulation, and an individualized herbal prescription that was regularly adjusted by a traditional Chinese herbalist. The standard formula contained 20 ingredients; the tailored prescriptions consisted of herbs chosen from a total of 81 herbs to match the unique pattern of Chinese diagnosis of each patient. To maintain concealment and the effectiveness of the placebo control, besides having a placebo designed to taste, smell, and look similar to the herbal formulas, all patients, after consulting a Chinese herbalist who was ignorant as to whether they were to receive real or dummy treatment, were required to complete a series of questionnaires and wait thirty minutes for the preparation of their herbs. The wait time was used to avoid patients identifying whether they were receiving standard, placebo, or customized formulations that take time to prepare.[27] Nonetheless a debate ensued. Questions were raised about whether the controls were truly inert and whether they could cause independent harmful or beneficial effects that might bias the outcome of the tria1.[28] Subsequently, the researcher identified a complex placebo containing 78.2% calcium hydrogen phosphate, 19.6% soy fiber, 0.3% cosmetic brown, 0.5% cosmetic yellow, 0.01% edicol blue, 0.09% identical licorice dry flavor, and 0.03% bitter flavor, and all parties agreed it was reasonable to conclude that the placebo was completely inert.[29] Other Chinese herbal trials have had to concoct similar complex placebos (e.g.[30]).

Homeopathy

At first glance, the least severe challenges for developing credible placebos would appear to exist for homeopathic research since their medications are often prescribed on lactose granules or pills which lend themselves to being easily disguised as placebos.[31] In fact many people claim that a homeopathic preparation versus a placebo control is little more than the comparison of two placebos.[32] In reality, the use of placebo in homeopathic research is probably better developed than for any other CAM modality, as witnessed by the fact that one meta-analysis was able to

locate 133 placebo-controlled trials in perhaps the most extensive meta-analysis yet conducted in CAM research for any individual therapy.[33]

This is not to say, however, that there are not daunting challenges even in this type of research. One problem arises from the nature of the modality itself, which often claims the need to select a unique remedy to match the particular configuration of the patient's complaints. Like Chinese medicine, the identical biomedical condition or disease may require treatment selected from many different homeopathic preparations because of the peculiar diagnostic system of homeopathy. For example two trials of homeopathy (one for diarrhea and one for diabetes) each selected from over 20 different homeopathic preparations—hence adding an untoward degree of complexity to the randomization and procedural aspects of the study.[31] This problem has been addressed quite creatively (and successfully) in a trial involving the treatment of fibromyalgia, in which only subjects who fit a specific symptom pattern are enrolled, hence the same medication could be prescribed within the confines of homeopathic theory.[34]

Acupuncture

The development and use of placebos in acupuncture trials has involved challenges due to both the nature of the modality and to controversies surrounding its hypothesized mechanism of action that exist within the discipline itself. Placebo acupuncture techniques, for example, involving the actual insertion of a needle at a non-indicated acupuncture point for the specific condition, have been criticized within the profession based upon the contention that insertion of an acupuncture needle anywhere near an indicated site always has the potential of producing an analgesic effect by releasing neurotransmitters which may effect pain thresholds.[35,36]

A general consensus has therefore developed that placebo acupuncture (i.e., involving the apparent insertion of a needle which does not actually break the skin) are superior controls in efficacy research. The second author's group[37] has successfully tested a sham technique that patients have difficulty distinguishing from true acupuncture, which involves the taping of a guide tube to the skin for both true and sham treatments. (In true acupuncture, the needle is actually inserted through the guide tube; in the sham groups the needle is tapped causing the guide tube to exert a slight pressure upon the skin.)

Recently, however, a potentially superior procedure (in the sense that the patient can actually observe the procedure being performed) has been developed independently by both German[38] and Korean-English[39] groups in which specially telescoping needles with blunt points are pressed against the skin, producing a distinct sensation at the point of contact to accompany the appearance of insertion. (The two needles appear basically the same, but differ in the way in which they are affixed to the skin.) The German placebo needle has already been used in one published randomized controlled trial, and demonstrated that it can successfully be used as a dummy control to show that acupuncture with penetration of the skin is more effective in an identical therapeutic setting, for the treatment of rotator cuff tendonitis.[40]

Healing Energies

Attempts to create placebo controls for CAM modalities that claim to heal with "vital energies" go back to at least the end of the eighteenth century with the famous experiments on mesmerism.[9] Many contemporary CAM therapies make similar claims to heal with analogous "subtle" energy.[41] To take one popular example, Reiki technique involves 13 hand positions (performed by a trained Reiki practitioner) on the fully clothed patient to theoretically reconnect the flow of the patient's universal life energy of life force through invisible channels within the body, the initial disruption of which is hypothesized to have resulted in (or is caused by) illness.

Creating a credible and reliable placebo to test Reiki claims was the goal of one research group. They tested an elaborate placebo condition involving the training of a group of sham practitioners in the correct hand motions by a Reiki master. (Because the demeanor of a true practitioner was deemed to be an important component of the therapy, however, these sham practitioners were also matched in terms of physique, warmth, and friendliness with the actual practitioners.) Real and sham practitioners were then tried out on a group of research volunteers, who basically could not distinguish between the two.[42]

While the results of the actual efficacy trial are not yet available, this genre of placebo has been employed for research involving other CAM modalities. All such sham interventions are potentially assailable within the CAM community, however, on the same basis as acupuncture placebos, namely that they may inadvertently contain an active component.

Chiropractic

Chiropractic manipulation, representing the most popular professional CAM modality in this country, employs a number of different therapeutic techniques, although the most widely employed involves spinal adjustment via the manipulation of the spine. This normally entails high velocity thrusts with either a long or short lever-arm, usua1ly with the specific objective of reducing pain and improving the patient's range of motion.[43] One systematic review of placebo-controlled trials, described several different techniques before concluding that sham (or placebo) procedures, while fairly rare in chiropractic research, were at least feasible.[44] Some of these included sham procedures involving (a) adjustments using minimal force,[45,46] (b) the use of an adjustable treatment bench allowing a sudden, non-manipulative change of position for the patient,[47] (c) a hands-on procedure which involved no manipulation at all,[48] and (d) the use of an adjusting instrument (activator) in which the force was set on "high" for the true chiropractic group but set on zero for the sham group.[49]

OTHER CONSIDERATIONS IN CAM PLACEBO TRIALS

Besides the specific challenges of particular CAM modalities there are some other general considerations in CAM placebo controls. These include: testing whether the placebo was concealed and the use of patient education as a placebo control.

Testing Placebo Blindness

Testing the credibility or the effectiveness of a placebo control is not a frequent question in clinical research.[50,51,52] Perhaps understandably, given the complex nature of placebo controls, it is not unusual, although far from universal, for CAM trials to test the effectiveness of their placebos. For example, in the acupuncture trials conducted by the second author and his colleagues, patients assigned to true vs. sham acupuncture conditions are routinely asked (usually following the first and final treatments) whether they think they received true acupuncture, sham acupuncture, or simply were unsure. Generally speaking we have obtained excellent blinding results for both our osteoarthritis and dental acupuncture trials as have the authors of a number of the placebo conditions discussed above. Given the complexities of developing placebo conditions for CAM research, however, coupled with the general skepticism surrounding the discipline among certain segments of the medical community, it is probably important that this simple question (or some variant thereof) always be asked in any CAM RCT in employing a placebo group.

Patient Education As A Control For Placebo Effect

While not a placebo in and of itself, patient education is sometimes included as an attention control for certain CAM trials based upon the theory that, while subjects assigned in such studies cannot be truly blinded, these groups are thought to partially "control for the placebo effect" by assuring similar amounts of contact with study personnel.[53] Our experience with this type of control indicates that the placebo effect is far too complex and variable to permit adequate interpretation of the data resulting from such trials.

Is There "Really" A Placebo Effect?

Recently, a meta-analysis of 114 RCTs which included both a placebo arm and a "no treatment arm" published in a high impact medical journal found little evidence that placebos have a powerful or significant clinical effect beyond natural history and regression to the mean.[54] This provocative study has already shifted conceptual and evidentiary discussion concerning the placebo effect. Nonetheless, many have noted that these authors took a deliberately narrow definition of placebo (the effect of a dummy intervention on a clinical outcome) and explicitly excluded the entire package of "non specific" effects embedded in the patient-physician relationship and clinical context that are often understood to be part of the placebo effect. These include such behaviors as: communication of concern, attention, monitoring and diagnostic procedures, labeling or explanation of disease, and more importantly, the impact of factors such as expectation, hope, and anxiety reduction.[55,56] Others have noted that evidence from many other research avenues indicate that a powerful placebo is a genuine clinical and research consideration.[57] However this new debate resolves, it should be noted that despite discounting a powerful placebo, the authors of this meta-analysis argued that placebo controls and a fastidious research approach are still necessary in order to control for bias.

Is There an "Enhanced" CAM Placebo?

While some have argued the absence of a placebo effect altogether, others have argued that CAM may have enhanced placebo effects.[58] Enhanced placebo effects refers to the notion that "two interventions may have different effects on patient outcome even though both [are] equivalent to placebo in clinical trials."[59] Evidence of an augmented placebo effect exists from multiple RCTs that compare sham devices (such as saline injections) to oral placebo (such as sugar pills).[60] Also several RCTs have recorded enhanced placebo effects in CAM trials.[58] Besides this evidence, the argument for enhanced CAM placebo effects rests on the fact that CAM has unique rituals and behaviors that have been associated with more powerful placebo outcomes. The first author of this paper is currently performing an RCT comparing the placebo effects of a CAM intervention with a mainstream intervention. (See reference 60 for details of the protocol.) Undoubtedly, much additional rigorous research remains to be performed to provide reliable evidence to support or reject such assertions. Whether such an effect actually exists has important consequences both for clinical practice and research.

ETHICAL CONSIDERATIONS IN PLACEBO-CONTROLLED TRIALS

From the very beginning of the RCT era of research the ethics of placebo-controlled trials has often been controversial.[61,62] This debate continues and has been especially activated by the World Medical Association's recent potential revisions of the Declaration of Helsinki which calls for testing of any new treatment to be done against the best current method where that exists, and not against a placebo.[63,64,65] Empirical evidence has shown that patients are denied treatment that could produce relief or prevent harm: a study of 18 placebo-controlled trials of ondansetron as prophylaxis against, or treatment for post-operative nausea and vomiting showed that around 2,620 patients had been denied existing drugs, which though not completely effective or without side effects do bring some relief.[66] Others have argued that placebo controls for such conditions as rheumatoid arthritis[67] and Alzheimer's disease[68] can no longer be considered because there is substantial evidence that existing drug treatments have both short-term and long-term benefits. In opposition, other researchers have argued for a more lenient allowance for placebo controls because the "harm and discomfort associated with the use of placebo controls are non-existent or are so small," and patients should be able to voluntarily give consent, and that placebo patients receive clinical attention and potential benefits from the placebo.[69]

Besides the benefit and harm argument, advocates of placebo control argue that placebo-controlled trials have more scientific validity and that this validity constitutes a fundamental ethical protection.[70] The scientific argument (related to the fastidious vs. pragmatic argument described earlier) continues to insist that a trial showing equivalence (or noninferiority) between new and previously established therapies is often not persuasive because such a trial relies on a historical assumption that is not provable and is not true. This assumption is that the trial, as conducted, has assay sensitivity, i.e, if there had been a placebo group present, the standard therapy could have been distinguished from placebo in that trial.[11] This assay sensitivity is a problem in the case of symptomatic treatments (e.g., pain, nausea) but also arises in end-point trials for many important drugs because many well-controlled trials were unable to demonstrate that any given end-point was statistically significantly improved compared to placebo (e.g., aspirin in some postinfarction trials).[26] Some argue that the values of patients and community

need to be included in the questions of placebo controls in order to sort through the complexity of risks and benefits.[71] (It should also be noted that occasionally, it has happened that a standard therapy is itself worse than no treatment.[72])

The problem of not allowing a placebo control also arises when one wishes to show that a new intervention has a positive effect without necessarily believing that it will be better than the standard therapy. Such a drug might be valuable in situations in which the standard cannot be used, even if it is less effective than the standard, but we need to know whether it is effective at all. A trial testing the new drug against the standard as a positive control may leave the question unanswered, while to use a placebo control in this situation would mean that some patients were not getting an active treatment when such a treatment was available. Of course, in some instance one could compare standard treatment plus new treatment (such as a CAM modality) versus just standard treatment. But this has the difficulty of a ceiling effect where the standard treatment already produces a significant result and makes any improvement in the new intervention less visible.

CONCLUSION

Placebo controls have been a central issue in CAM research since the beginning of alternative medicine. Placebo controls are a critical methodological safeguard in research. The issues of placebo controls are necessarily complex and clinicians and researchers need to be aware of the significance of the options.

REFERENCES

1. Kaptchuk TJ, Eisenberg DM. Varieties of healing: 1: Medical pluralism in the United States. Ann Inter Med 2001; 135:189-195.

2. Leland PW, Empiricism and its causes. Boston Med Sur J 1852; 47:283-295.

3. Conferences on Therapy. NY J Med 1946; 46:1718-1727.

4. Joyce CRB. Placebo and complementary medicine. Lancet 1994; 344:1279-1281.

5. Lynoe N. Is the effect of alternative medical treatment only a placebo effect? Scand J Sac Med 1990; 18:149-153.

6. Brown W A. The placebo effect. Sci Amer 1998; 68-73.

7. Spiro H. The Power of Hope. New Haven: Yale University Press, 1998.

8. Healy D. The Anti-Depressant Era. Cambridge: Harvard University Press, 1999.

9. Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. Bull Hist Med 1998; 72:389-433.

10. Lilienfeld AM. Ceteris paribus: the evolution of the clinical trial. Bull Hist Med 1982; 56:1-18.

11. Temple RJ. When are clinical trials of a given agent vs. placebo no longer appropriate or feasible. Cont Clin Trials 1997; 38:613-620.

12. Kaptchuk TJ, Edwards RA, Eisenberg DM. Complementary medicine: efficacy beyond the placebo effect. In: Ernst E (ed) Complementary Medicine: An Objective Appraisal. Oxford: Butterworth Heinemann, 1996.

13. Sullivan MD. Placebo controls and epistemic control in orthodox medicine. J Med Phil 1993; 18:213-231

14. Lyon ML. Order and healing: the concept of order and its importance in the conceptualization of healing. Med Anthropology 1990; 12:249-268.

15. Kaptchuk TJ. Powerful placebo: the dark side of the randomized controlled trial. Lancet 1998; 351:1722-1725.

16. Van Weel C. Examination of context of medicine. Lancet 2001; 357: 733.

17. Feinstein AR. Clinical Epidemiology Philadelphia: W.B. Saunders, 1985.

18. Schwartz D, Lellouch J. Expanatory and pragmatic attitudes in therapeutic trials. J Chron Dis 1967; 20:637-648.

19. Black N. Why we need observational studies to evaluate the effectiveness of health care. Brit J Med 1996; 312:1215-1218.

20. Uhlenhuth EH, Ricels K, Fisher S, Park LC, Upman RS, Mock J. Drug, doctor's verbal attitude and clinic setting in the symptomatic response to pharmacology. Psychopharmacologia 1966; 9:392-418.

21. Kaptchuk TJ. The double-blind, randomized placebo-controlled trial: gold standard or golden calf? J Clin Epidem 2001; 54:541-549.

22. Rochon PA, Binns MA, Litner JA, Litner GM, Fischbach MS, Eisenberg DM, Kaptchuk TJ, Stason WB, Chalmers TC. Are randomized control trial outcomes influences by the inclusion ofa placebo group?: A systematic review of nonsteroidal anti-inflammatory drug trials for arthritis treatment. J Clin Epidem 1999; 52:113-122.

23. Spilker B. Guide to Clinical Trials. New York: Raven Press, 1991.

24. Gotzsche PC. Is there logic in the placebo? Lancet 1994; ii:925-926.

25. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. Brit J Med 1996; 313:36-39.

26. Temple R, Ellenberg 55. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. Ann Intern Med 2000; 133:455-463.

27. Bensoussan A, Talley NJ, Hing M, Menzies R, Guo A, Ngu M. Treatment of irritable bowel syndrome with Chinese herbal medicine: a randomized controlled trial. J Amer Med Assoc 1998; 280:1585-1589.

28. Kaptchuk TJ. Chinese herbal medicine for irritable bowel syndrome. [Letter] J Amer Med Assoc 1998; 282:1035-1036.

29. Bensoussan A, Talley NJ, Menzies R. Chinese herbal medicine for irritable bowel syndrome. [Response to letter] J Amer Med Assoc 1998; 282:1035-1036.

30. Sheehan P, Atherton DJ. A controlled trial of traditional Chinese medicinal plants in widespread non-exudative atropic eczema. Brit J Dermatology 1992; 126:179-184.

31. Jacobs J. President's message: homeopathic research with heart. J Amer Inst Homeopathy 2000; 93:121-124.

32. Vandenbroucke JP, de Craen AJM. Alternative medicine: a “mirror image” for scientific reasoning in conventional medicine. Ann Intern Med 2001; 135:507-513.

33. Linde K, Clausius N, Ramirez G, et al. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. Lancet 1997; 350:834-843.

34. Fisher P, Greenwood A, Huskisson EC, Turner P, Belon P. Effect of homeopathic treatment on fibrositis (primary fibromyalgia). Br Med 1989; 299:365-366.

35. Vincent C, Lewith G. Placebo controls for acupuncture studies. J R Soc Med 1995; 88:199-202.

36. Kaptchuk TJ. Acupuncture: theory, efficacy and practice. Ann Intern Med 2002; 36:374-383.

37. Berman BM, Singh BB, Lao L, et al. A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. Rheumatology 1999; 38:346-354.

38. Streitberg K, Kleinhenz J. Introducing a placebo needle into acupuncture research. Lancet 1998; 352:364-365.

39. Park J, White AR, Ernst E. New sham method in auricular acupuncture. [Letter] Arch Intern Med 2001; 161:894.

40. Kleinhenz J, Streitberg K, Windeler J, Gubbacher A, Mavridis G, Martin E. Randomised clinical trial comparing the effects of acupuncture and a newly designed placebo needle in rotator cuff tendonitis. Pain 1999; 83:235-241.

41. Kaptchuk TJ. History of vitalism: In: Micozzi MS. Ed. Fundamentals of Complementary and Alternative Medicine. 2nd ed. New York: Churchill Livingstone, 2001:43-56.

42. Mansour AA, Beuche M, Laning G, Leis A, and Nurse J. A study to test the effectiveness of placebo Reiki standardization procedures developed for a planned Reiki efficacy study. J Alt Med Comp Med 1999; 5:153-164.

43. Kaptchuk TJ, Eisenberg DM. Chiropractic: origins, controversies, and contributions. Arch Intern Med 1998; 158:2215-2224.

44. Ernst and Harkness 2001; in press.

45. Waagen GN, Haldeman S, Cook G, et al. Short term trial of chiropractic adjustments for the relief of chronic low back pain. Manual Med 1986; 2:63-67.

46. Triano JJ, McGregor M, Hondras MA, Brennan PC. Manipulative therapy versus education programs in chronic low back pain. Spine 1995; 20:948-955.

47. Nielsen NH, Bronfort G, Bendix T, et al. Chronic asthma and chiropractic spinal manipulation: a randomized clinical trial. Clin Exp Allergy 1995; 25:80-88.

48. Balon J, Aker PD, Crowther ER, et al. A comparison of active simulated chiropractic manipulation as adjunctive treatment for childhood asthma. New Engl J Med 1998; 339:1013-1020.

49. Angler MJ, Kelin RGT, Dorman TA, et al. A new approach to the treatment of chronic low back pain. Lancet 1987;2:143-147.

50. Hughes JR, Krahn D. Blindness and the validity of the double-blind procedure. J Gin Psychopharm 1985;5:138-142.

51. Margraf J, Ehlers A, Roth WT et al. How "blind" are double-blind studies? J Consult Clinc Psychol 1991;59:184-187.

52. Basoglu M, Marks I, Livanou M, Swinson R. Double-blindness procedures, rater blindness and rating of outcome. Observations from a controlled trial. Arch Gen Psych 1997;54:744-748.

53. Lorig K, Lubeck D, Kraines RG, Seleznick M, Holman HR. Outcomes of self-help education for patients with arthritis. Arthritis and Rheum 1985; 28:7-12 (Appendix 2).

54. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Eng J Med 2001; 344:1594-1602.

55. Greene PJ, Wayne PM, Kerr CE, Weiger WA, Jacobson E, Goldman P, Kaptchuk TJ. The powerful placebo: doubting the doubters. Advances in Mind-Body Medicine 2001; 17:298-307.

56. Hrobjartsson A. What are the main methodological problems in the estimation of placebo effects? Clin Epidemiol 2002; 55:430-435.

57. Kaptchuk TJ. [Letter] N Eng J Med 2001; in press.

58. Kaptchuk TJ. The placebo effect in alternative medicine: can the performance of a healing ritual have clinical significance? Ann Intern Med 2002; 136:817-825.

59. Vickers AJ, de Craen AJM. Why use placebos in clinical trials? A narrative review of the methodological literature. J Clin epidemiol 2000; 53:157-161.

60. Kaptchuk TJ, Goldman P, Stone DA, Stason WB. Do medical devices have enhanced placebo effects? J Clin Epidem 2000; 53:786-792.

61. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. N Eng J Med 1994; 331:494-397.

62. De Deyn PP, D'Hooge RD. Placebos in clinical practice and research. J Med Ethics 1996; 22:140-146.

63. Christie B. Doctors revise Declaration of Helsinki. Brit Med J 2000; 321:913.

64. Rothman KJ, Michels KB, Baum M. Declaration of Helsinki should be strengthened. Brit Med J 2000; 321:442-445.

65. Ferriman A. World Medical Association clarifies rules on placebo controlled trials. Brit Med J 2001; 323:825.

66. Aspinal RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. Brit Med J 1995; 311:844-846.

67. Stein MC, Pincus T. Placebo-controlled studies in rheumatoid arthritis: ethical issues. Lancet 1999; 353:400-403.

68. Knopman D, Kahn J, Miles S. Clinical research designs for emerging treatments in Alzheimer disease. Arch Neurol 1998; 55:1425-1429.

69. Emanuel EJ, Miller FG. The ethics of placebo-controlled trials—a middle ground. N Eng J Med 2001; 345:915-919.

70. Emanuel EJ, Wendler D. Grady C. What makes clinical research ethical? J Amer Med Assoc 2000; 283:2701-2711.

71. Karlawish JHT, Whitehouse PJ. Is the placebo control obsolete in a world after donepezil and vitamin E? Arch Neurol 1998; 55:1420-1424.

72. Gilbert JP, McPeek B, Mosteller F. Progress in surgery and anesthesia: benefits and risks of innovative therapy. In Bunker JP, Barnes BA, Mosteller F (eds) Costs, Risks and Benefits in Surgery. New York: Oxford University Press, 1977.

COMPLEMENTARY AND ALTERNATIVE MEDICINE (CAM) CLINICAL TRIALS FOR SYMPTOMS: DEVELOPMENT OF APPROPRIATE INTERVENTIONS

Lonnie Zeltzer, MD*
Professor of Pediatrics, Anesthesiology, Psychiatry and Biobehavioral Sciences
Director, Pediatric Pain Program
Associate Director, Patients and Survivors Program
Associate Director, Integrative Oncology Program
Jonsson Comprehensive Cancer Center
David Geffen School of Medicine at UCLA

Jonathan Bergman, BS
Second Year Student
David Geffen School of Medicine at UCLA

**Corresponding author:*
Lonnie Zeltzer, MD
UCLA School of Medicine
22-464 MDCC, 10833 LeConte Avenue
Los Angeles, CA 90095-1752
Tel: (310) 825-0731
Fax: (310) 794-2104
Email: lzeltzer@mednet.ucla.edu

COMPLEMENTARY AND ALTERNATIVE MEDICINE (CAM) CLINICAL TRIALS FOR SYMPTOMS: DEVELOPMENT OF APPROPRIATE INTERVENTIONS

ABSTRACT

Purpose:

This paper is intended to review relevant issues related to individualization versus standardization in complementary and alternative medicine (CAM) clinical trials, to use several CAM interventions to illustrate both approaches, and to provide design recommendations to facilitate a more unified approach for reviewers who are asked to critique CAM studies.

Methods:

Building on suggestions made during special NCI review panel on CAM, literature relevant to CAM study design was reviewed, as were CAM clinical trials for different CAM interventions.

Results:

Both standardized and individualized interventions in CAM clinical trials have been used to successfully evaluate the efficacy of CAM interventions. Optimal study design varied by type of CAM intervention.

Conclusion:

The schism between Western and CAM beliefs regarding health, medical interventions, and study design must be bridged to maximize the utilization of CAM within Western biomedical frameworks. The optimal approach to the issue of standardized versus individualized CAM interventions in clinical trials appears to be an integrative one that develops a strategy with some room for individualization within that strategy. The future directions of CAM research will necessitate a shared learning experience of CAM clinicians, biomedical researchers, and social scientists.

Key Words:

Complementary and alternative medicine, individualization, clinical trials, symptoms, standardization, study design.

INTRODUCTION

While complementary and alternative medicine (CAM) has been practiced throughout the world for hundreds of years, its utilization in the developed world has increased sharply over the past two decades. A recent epidemiological study on CAM usage reported that approximately one-third of the population in both the United States and in the United Kingdom visits CAM practitioners.[1,2] More than two-thirds of family practitioners in the United States have expressed an interest in increasing their utilization of CAM therapies.[3] As the clinical interest has increased, both the scientific and medical practice communities have called for scientific evidence documenting the efficacy of the CAM practices that the public seeks and uses. The first documented CAM trials that used blinding and placebo date back to 1885.[4] However, the paucity of published CAM clinical trials that utilized the same types of scientific design and methodologies as pharmaceutically derived drug trials or other more traditional non-drug interventions has left the scientific and physician communities with doubts about the value of CAM interventions as more effective than placebo. On the other hand, many CAM practitioners believe that what they do works, and they have difficulty understanding the scientific and traditional medical community's need to impose scientific constraints to "prove" by other types of standards how, why, when, and if the various CAM interventions work. This seeming schism between the scientific and medical practice communities and CAM practitioners appears, at least in part, to be rooted in different paradigms about healing.

Western science and medical practice tend to be reductionistic, focusing on determining the "truth" about an entity or phenomenon and testing what works "best." There is an emphasis on distinction and dividing entities into their component parts and determining which part or what hierarchy of parts is the most effective component. In scientific clinical trials testing an intervention, whether drug or non-drug, the typical strategy is first to test for toxicity, then the potential of efficacy through testing likely effect size, and then finally to test for efficacy in a large, double-blind, randomized, controlled study.

On the other hand, most CAM traditions tend to be holistic, presuming that mechanisms of healing relate to linkages and systems, rather than specific component parts. Characterizing this "whole" can be more complicated than describing "parts" in a reductionistic health and illness paradigm. Since systems and relationships can and do change, holistic (e.g., CAM) practitioners see their task as helping a system to find balance, efficiency, or optimal homeostasis rather than searching for the "truth" (e.g., correcting the defective part to achieve wellness). This difference in emphasis of the intervention in the two systems also creates differences in the outcome targets, with significant differences in the two systems in their characterization of "progress" and change. CAM practitioners are looking for points of balance and often work in the present, such as initiating the flow of "Qi" (pronounced "chee") to reduce "heat" in an organ, for example, or enhancing "prana" (also an energy flow or "life force") to balance a "chakra" (often described as a focal point of energy generation). While symptoms are expected to improve over time within this paradigm and wellness to be enhanced, the time points for achieving endpoints are more elusive and it is rather the balance and working of the system as a whole that is examined. In a more traditional medical practice system, the goals tend to be more specific and future-oriented. That is, the expectation in comparing Treatment A to Treatment B or to a placebo is that, over a specified period of time, Treatment A will be more effective than Treatment B or than the

placebo as defined by changes in certain specific measured biological and/or symptom parameters.

Another difference between the two paradigms is the focus on the responsibility for wellness. In most CAM practices, the intervention is aimed at optimizing the patient's own abilities to achieve well-being. In fact, in some CAM practices, such as in yoga, the "patient" is called the "student" and the "teacher" guides the student in achieving physical, emotional, cognitive, and spiritual harmony and balance, but it is ultimately up to the student to want to work at achieving this balance. In most traditional clinical medicine, there is a hierarchy of responsibility for the patient achieving wellness, with the physician often assuming that responsibility by prescribing the best treatment for the patient based on what the physician believes is the "cause" of the problem. While the patient is relied upon to adhere to the treatment regimen prescribed by the physician, rarely are the patient's beliefs about the likely efficacy of treatment considered and the patient's role in the path to wellness from a mind-body connectivity perspective is often given lower valence. Table 1 provides a summary of the highlights of the practice of CAM interventions.[3,5-6]

Table 1. General Strategy of CAM Interventions

- The practitioner typically makes an individualized diagnosis and develops a treatment approach based on that diagnosis, rather than using a "standard" approach based on the "category" of diagnosis (e.g., insulin for diabetes).
- Evaluation is an ongoing process, so the treatment may change over time, depending upon the patient's response and changes found in the ongoing evaluation process.
- The overall goal of most CAM approaches is to maximize the body's inherent healing ability and restore "balance" or homeostasis.
- While there may be specific symptoms that bring the patient to the CAM clinician, the treatment is organized around the patient as a whole, including an integration of the individual's physical, emotional, cognitive, and spiritual health.
- This "whole person" orientation is different from the more traditional Western biomedical focus on diagnosis of a pathogenic process and then application of what is known to treat that "disease."

Within the framework of some of the clinical practice differences in the two healthcare paradigms described above, the overlay of research design and methodology in establishing clinical trials on top of clinical practice in order to provide scientific evidence of efficacy of the various CAM practices creates its own set of problems. Given the many differences in CAM and traditional medical paradigms, it is no wonder that the development of research to guide the accepted practice of CAM interventions and to facilitate the integration of CAM and traditional therapies has been slow. This seeming clash comes to a head for scientific reviewers who carry out their roles as members of study sections trying to critique grant applications for CAM

studies. Reviewer inconsistency because of lack of clear guidelines often creates frustration for investigators. Thus, there may be a heterogenicity of the quality of the design of what studies get funded. This paper is embedded within a larger body of discussion and review on CAM clinical trials as part of a special review panel set up by the CAM Division of the NCI. Thus, this paper will address only a component of the many challenges facing both investigators of CAM interventions as well as the grant application reviewers of these proposed CAM clinical trials. The intent of this paper is to discuss the pros and cons of one of the major design issues in CAM clinical trials: individualized versus standardized interventions. The paper will review arguments for each approach and then will use a number of CAM intervention practices to illustrate through specific studies both approaches. Finally, recommendations will be provided for design of CAM intervention clinical trials, with the goal of enhancing a more unified approach for reviewers to consider in critiquing CAM studies.

STANDARDIZED VS. INDIVIDUALIZED APPROACHES

Nahin and Straus have suggested three major approaches to studying CAM Interventions.[7] The first approach is to study a whole intact CAM system for a specific disease or problem. One example is comparing traditional Chinese medicine to a naturopathic system or to a conventional treatment approach for temporomandibular joint pain. Another approach is to examine a component part adapted from a CAM system, such as acupuncture, compared to a placebo acupuncture procedure or other treatment for a specific problem. The third approach is to test a single standardized intervention, such as testing an herbal treatment (e.g., hypericum) for a conventionally diagnosed disease (e.g., depression). Within these three approaches, especially the first two, there are two major study design issues to consider: 1) a standardized intervention vs. an individualized approach, and 2) a trial of a single intervention that does not accurately reflect true clinical practice vs. a multifaceted intervention trial that is complicated to design and implement.

Despite an emphasis on multimodality treatment of the patient as a whole, many CAM studies have examined only one component taken from a whole treatment system (e.g., acupuncture needling, without addition of other components of TCM such as Chinese herbs, dietary changes, physical components like Qi Gong, moxibustion, etc.). Often this approach is carried out in funded research because it appears easier from the outside to design and seems more familiar to most clinical scientists, since the study has a standardized specific intervention that is compared to a placebo in a randomized, double blind, controlled clinical trial. However, this seeming simplicity begins to unravel as the design takes shape and as the larger heuristic value of the study is examined from a practical standpoint. That is, will the study actually test the CAM modality as it is practiced in the CAM community? This is the issue of testing external validity of the whole system (e.g., TCM) versus component part (e.g., acupuncture). However, even if there is validity in testing a component of a system, issues related to the implementation of a single intervention differ across different CAM therapies. In yoga, for example, there is a difference in the practice of a beginning student versus an experienced student, and as the student develops awareness and understanding, the actual practice of yoga changes over time. Differences in physical ability, and changing effects of surgery, radiation, or chemotherapy need to be considered in the development and changes in the yoga practice. Psychological status and emotional well-being have to be considered in the yoga practice and can change situationally; in the yoga example, all of the above factors would need to be considered in determining which set

of poses the intervention would comprise. Table 2 reviews some pros and cons of considering an individualized versus a standardized intervention within a study design that tests a whole CAM system, while Table 3 reviews pros and cons for a design that tests a component part.

Table 2. Study of CAM System as a Whole: Individualization of Intervention

Pro
• Can most accurately reflect how the CAM intervention is practiced in the community and thus addresses external validity.
• May maximize main effects in a shorter amount of time than use of components of the CAM intervention.
• If effect sizes are large, then sample sizes can be relatively smaller.
• More feasible if a single Western medicine diagnosed entity is the focus of study (e.g., osteoarthritis). To be most effective, the diagnosis or problem according to the CAM practice needs to be defined and included in study entry criteria so that outcome can be analyzed within the framework of the CAM diagnosis.
• Can be compared to other CAM systems or standard treatment, with side effect profile and cost effectiveness considered in outcomes.
• More likely to get cooperation from CAM clinician community which is often frustrated with research that takes a CAM modality out of context.
Con
• Difficult to design appropriate controls.
• Need a series of dismantling studies to determine whether the package as a whole is needed or whether there are critical components and which of those components is/are responsible for the outcome.
• May be difficult for CAM clinicians to agree upon all the different components: more decisions go into the description and adherence to the treatment.

Table 3. Study of CAM Component: Individualization of Intervention

Pro

- Can most accurately reflect how the CAM modality is practiced in the community.
- CAM modality may produce a greater effect because it is geared to the individual patient and embedded in the intent of the CAM practice.
- If effect sizes are large, then sample sizes can be relatively smaller.
- Can maximize on effect size and then do next-step study comparing the individualized approach to a standardized approach (e.g., individualized vs. standardized herbal treatment for irritable bowel syndrome).[24]
- Optimized by including the CAM based diagnosis and examining outcomes in relation to both the Western and CAM diagnosis.

Con

- A standardized approach makes the design of a control group easier. A standardized approach makes the design of a double blind study more feasible.
- May be simpler to start with a controlled, standardized approach and, if effective, then do a next-step comparison with an individualized approach to treatment.
- A standardized approach allows more readily for internal validity and generalizability of the intervention.
- A standardized approach is easier for minimizing clinician differences (e.g., training and experience of the clinician would be expected to have a greater impact on individualized treatments).
- Group interventions are more amenable to a standardized approach (e.g., a hypnotherapy group).
- Underlying mechanisms of action are more easily assessed in a standardized approach.

CAM STUDY DESIGN CONFOUNDS

Standardizing CAM therapies represents a distinct challenge, since CAM treatments are centered around differential therapeutics (previously referred to as individualization) designed for each patient. Whereas biomedical models categorize a diagnosis and apply an appropriate treatment, CAM practitioners analyze each individual as a whole, evaluate the patient continuously, and adjust therapy over time, focusing on symptoms rather than primary biomedical pathology.[8,9] As

noted, the goal of treatment is to maximize the body's inherent healing ability and restore homeostasis; to that end, treatment is individualized in regard to the length as well as nature of the treatment. Despite this obstacle, the challenge of analyzing CAM therapies in a manner acceptable to biomedical practitioners has not proved insurmountable. One study following patients treated at a CAM clinic found that 58/107 patients with a given biomedical diagnosis were treated with the same class of CAM therapies, suggesting that analyzing CAM treatments within both biomedical and CAM paradigms is possible.[10] However, Mercer, et al. emphasized the difficulties in standardizing CAM treatments and suggested that getting a group of CAM practitioners to agree and adhere to the same standardized treatment would pose challenges.[6] There would also be differences between CAM practitioners and medical scientists in conflicting concepts and theories of health and disease, lack of agreement about diagnostic criteria, contrasting views about the therapeutic process, and different theories of causation.[6]

When pooling data, some authors have struggled with the question of whether or not the concentration of the remedy (dosing regimen) used and the diagnostic category (condition to which it is being applied) were comparable.[3] Others have addressed this concern by comparing the general efficacy of CAM treatments to placebo.[11,12,13] While such approaches can be employed for individual trials, they complicate the design of meta-analyses. Blinding to treatment condition is another challenge to the design of CAM clinical trials. Some believe that, since the CAM clinician is an integral part of the intervention, blinding the clinician to treatment condition is impossible.[5] Blinding on the patient side could be designed, but double-blind trials might not be believable. While some authors have supported the notion of randomization without blinding,[14] others have maintained that double-blinding is essential for concrete findings that establish a causal relationship.[10] One solution proposed suggested asking the patients to which group they believed they had been randomized during the trial; this would help assess the efficacy of the blinding process, at least for patients.[10] In any case, the outcome data should be obtained by observers who are blind to treatment condition and quality checks could be made on this also. Additionally, many authors have expressed a fear that both the control as well as the intervention group would seek other CAM treatments in addition to the one being studied, rendering a finding of causation impossible.

Despite these challenges, a number of randomized, double-blind studies have been carried out within the individualized framework of CAM practices. One group conducted a randomized, double-blind, placebo-controlled trial of 60 patients with persistent mild traumatic brain injury. Homeopathic treatments were designed for each patient, then the subjects received either CAM treatment or placebo, with clinically and statistically significant improvement seen for the CAM treatment group compared to the control group.[11] Other authors have used similar designs, wherein a CAM therapy was established for each patient, a control system was prepared, and the results of CAM therapy versus placebo were reported; this design has been used to study acute otitis media in children[12] as well as acute childhood diarrhea.[13] As a report exploring the issues regarding the use of randomized control trials in complementary therapies explained, differential therapeutics is possible *within* groups randomized to CAM therapies.[14] The CONSORT statement regarding the improvement of the quality of reporting of randomized controlled trials maintained that the development of specific primary and secondary outcome measures was essential if trial results were to be trusted.[15] Although the evaluation of primary and secondary outcome measures in CAM studies is challenging, sound studies can be designed by integration

of biomedical and CAM models, patient and practitioner evaluations, CAM clinician and biomedical physician assessments, and quantitative as well as qualitative data.[14,16-18]

Several authors worry that the description of allocation concealment and the reporting of drop-outs and withdrawals is less stringent in CAM trials compared to biomedical trials.[14,19] This belief represents a challenge to the CAM community as well as to the biomedical community. The need to educate the biomedical community in the tenets of CAM therapy and the CAM community in the biomedical research model is vital to the advancement of the validity of both paradigms. Table 4 summarizes study design confounds in considering CAM clinical trials.

For readers concerned about the validity of randomized CAM clinical trials, instruments have been developed to assess the quality of such reports. The scale most widely used to evaluate CAM therapies is the Jadad scale, which assesses the completeness of reporting using three items that assess random allocation, double-blinding, and the reporting of dropouts and withdrawals.[20] Jadad's scales are useful to assess discrimination, face validity, and reliability. Once good CAM trials are chosen by Jadad's scale, results appear to be as reliable as well-designed biomedical studies.[21]

Table 4. CAM Study Design Confounds

- Large variability in the ways each CAM treatment is practiced (e.g., different needling techniques in acupuncture; different types of yoga such as Iyengar, Hatha, Bikram, etc.).
- Within approaches, treatments may vary for patients presenting with the same biomedical diagnosis because CAM practitioners focus on the symptoms rather than on a specific pathology.
- The number and length of treatments, and the specific treatments used, may vary both between patients and for an individual patient during the course of treatment (e.g., in acupuncture: selection of points, depth of needle insertion, angulation of needles, use of electro stimulation, movement of needles, duration of needling, need to achieve "Qi," and frequency and scheduling of treatments).
- Need for assessment of the use of other CAM interventions or "cheating" by the control or intervention group, since CAM therapies are easier to obtain than more traditional biomedical therapies.
- Placebo effect is enhanced because CAM treatments typically involve extended and intensive interactions between the patient and practitioner, and thus the difference between active treatment and placebo effects may be minimized.

STANDARDIZED VS. INDIVIDUALIZED APPROACHES: EXAMPLES WITHIN CAM INTERVENTIONS

Because the issues related to individualization or standardization of CAM interventions may differ across CAM treatments, the following section will provide examples of these issues as they relate to different categories of CAM interventions in clinical trials.

Traditional Chinese Medicine and Acupuncture

The issues of individualization versus standardization of intervention appear to have been explored more thoroughly in Traditional Chinese Medicine (TCM) and acupuncture studies than in most studies of other CAM therapies. Review of this literature found that standardized intervention trials were often performed allowing some degree of individualization within the treatment arm. In a multicenter, double-blind, placebo-controlled study of childhood bronchial asthma in Taiwan, three different groups of Chinese herbs were compared to placebo. In this study, the individualization of herbal treatment (the CAM intervention) was based on the TCM diagnosis (all subjects had asthma by physician diagnosis). A team of CAM practitioners used TCM criteria to classify 310 asthmatic children according to three diagnoses: deficiency of kidney energy, deficiency of spleen energy, and deficiency of kidney and spleen energies. Each of the three CAM diagnoses was treated either with its remedy herb or with placebo, and clinical efficacy was assessed by allergists, TCM practitioners, and parents. Although a placebo effect was seen, significant improvement was noted in all three treatment arms relative to controls.[22] In a prospective, randomized, controlled, double-blind clinical trial testing the effectiveness of acupuncture treatment at real points vs. sham points in 67 patients with osteoarthritis of the hip, the benefit of individualization was assessed. Outcome parameters included pain as assessed by VAS, functional impairment, activity in daily life, and overall satisfaction before and after treatment. Significant improvement was found in both groups compared to previously studied untreated populations, but no differences were seen between the treatment and control (sham acupuncture) groups. Based on their findings, the authors questioned the benefit of individualizing therapy for osteoarthritic hip patients.[23] A 1998 randomized, double-blind, placebo-controlled study of Chinese herbs by Bensoussan, et al. compared individualized treatment, standardized treatment, and placebo in 116 patients with irritable bowel syndrome (IBS). Subjects were randomly assigned to one of the three groups and evaluated by a traditional Chinese herbalist as well as by a gastroenterologist. Standardized and individualized treatments both led to significant improvement compared to control immediately after treatment, but improvement after 14 months was seen only in the individualized treatment group.[24] Reviewing the TCM literature, Critchley, et al.[8] concluded that the development of evidence-based TCM was attainable using a traditional scientific method with individualization of treatment by TCM practitioners. However, in herbal studies there is a need to allow for specification of the contents, since there could be variation across herbal batches. There need to be independent observers as well as randomization to treatment and placebo arms, and detailed recording of treatments is needed to allow results to be reproduced.[8] An excellent discussion of the key issues in designing acupuncture trials can be found in Lao, et al.[25]

Yoga

Yoga treatments are rarely described based on their individualization for specific patients, but rather the type of yoga is often the unit of description (e.g., Iyengar, Bikram, Hatha, etc.). Garfinkel, et al.[26] reported on a randomized, single-blind, controlled trial of yoga in 42 carpal tunnel syndrome patients. Eleven yoga postures were specifically designed for strengthening, stretching, and balancing each joint in the upper body, along with relaxation. All subjects in the treatment group received the same sets of poses, with postures adjusted to meet the physical conditions of the students. Subjects treated for eight weeks were compared to patients who were offered splints to supplement their current treatment, and the yoga group had significantly increased grip strength and Phalen sign, as well as decreased pain, compared to the controls.[26] In a randomized, double-blind, placebo-controlled study of 18 mild asthma patients, subjects were randomized to yoga versus a control condition. The yoga included pranayama slow deep breathing exercises for 15 minutes twice a day for two consecutive 2-week periods. Subjects in the treatment arm breathed through a device that imposed 1:2 inspiration:expiration breathing, equivalent to pranayama breathing, while control subjects breathed through a matched placebo device that did not impose a set type of breathing. The pranayama breathing group required a significantly increased dose of histamine to provoke a 20% reduction in FEV1 compared to the control group.[27] A prospective, randomized, controlled trial employing standardized treatment for coronary artery disease (CAD) included 42 men with angiographically proven CAD who were randomized to control or yoga intervention and followed for one year. The yoga group spent four days at a yoga residential center, after which they carried out yoga exercises at home for an average of 90 minutes per day for one year. During that time, they visited a yoga center once every two weeks for monitoring and evaluation, including compliance. Both the treatment and control groups received recommendations for control of risk factors, diet control, and aerobic exercise. Patients in the treatment group suffered less angina, had decreased body weight, lower LDL, serum cholesterol, and triglycerides, fewer coronary artery bypass graft procedures, and slower disease progression.

Energy Therapies, Including Reiki

In a single-blind, three-group study of post-operative pain, 108 patients were randomly assigned to receive either therapeutic touch, a placebo control intervention mimicking therapeutic touch, or standard intervention with opioids. For subjects in the treatment arm, a therapeutic touch nurse assumed a meditative state of awareness, then moved her hands over the patient in response to areas of accumulated tension, an individualized intervention. For subjects in the sham intervention group, the same hand gestures were used with no attempt to assume a meditative state of awareness or to respond to patient energy fields. Although pain, as assessed by VAS, was not significantly decreased in the therapeutic group relative to controls, the therapeutic touch group required less analgesia.[29] Similarly, a single-blind trial of therapeutic touch versus sham touch for pain relief in 99 burn victims used an individualized approach based on patients' responses to energy fields. Therapeutic touch (TT) was provided once a day for five days and the TT practitioner assessed energy fields and implemented clearing, directing, or energy flow-balancing techniques in response to the fields emitted by the patient. Treatment varied in duration based on the practitioner's judgment and the state of the subject's energy field. Subjects who received TT reported decreased pain, anxiety, and total CD8+ lymphocytes compared to the sham control group.[30] Many energy practitioners would argue that creating a sham energy

intervention is impossible, since all people emit energy fields, even if they are not consciously aware of this.

Hypnotherapy

In a study that utilized a structured hypnotic strategy but individualized specific imagery to patients' interests and experiences, hypnosis was compared to nonhypnotic relaxation and distraction techniques for reducing pain and anxiety in 33 pediatric cancer patients during bone marrow aspirations or lumbar punctures. The hypnosis group received a structured hypnotic protocol with individualized imagery, while the control group was instructed on deep breathing, distraction, and practice sessions to reduce anxiety. Hypnosis was shown to be more effective than non-hypnotic techniques for the reduction of pain and anxiety.[31] Two different standardized hypnotherapy strategies were compared to a control condition in a randomized study of weight loss in 60 obese patients with obstructive sleep apnea receiving nasal continuous positive-airway-pressure treatment. Both hypnotherapy groups showed increased weight loss relative to controls.[32] Another design strategy is to standardize the treatment but stratify the patients by hypnotizability. One trial of 39 patients with moderate asthma utilized a single hypnotic technique to treat 39 adults, with relief seen in patients who were moderately or highly hypnotizable.[33] A smoking cessation study of 87 volunteers randomized to a hypnosis, relaxation, or wait-list control group showed benefit of hypnotherapy only for individuals able to enter medium or deep hypnotic states.[34] A study of chemotherapy-related nausea and vomiting compared a standardized hypnotic strategy with individualization according to patient needs, a relaxation/distraction intervention, and a standard treatment control condition in 51 children with cancer. While significantly greater symptom reduction was found in both treatment conditions, the hypnosis group was significantly more effective than the comparison intervention.[35]

Massage Therapy and Other Body Work

In a study of foot massage to reduce pain and nausea in cancer patients, 87 subjects underwent massage on two occasions and acted as their own control on a third; the order of massage versus control was counterbalanced. Massages were standardized without individualization to each patient. Massage was found to improve pain, nausea, and relaxation relative to the control period.[36] A prospective, randomized, controlled, single-blind study of 78 patients undergoing diagnostic cardiac catheterization included a standardized 10-minute massage, while control subjects spent 10 minutes of quiet time with a massage therapist; the massage group showed decreased anxiety scores.[37] Standardization in some studies has been aided by massage machines. In one study, 25 patients undergoing a colectomy underwent active mechanical massage by intermittent negative pressure on the abdominal wall for seven days after preoperative randomization. The computer-driven massage device allowed a reproducible massage that was similar in nature to a manual massage. Patients in the treatment arm used fewer analgesics and reported less pain than those in the control group.[38] In a study of cardiac vagal tone using myofacial trigger-point massage in 30 healthy subjects, individualization was used for location of trigger points, after which standardized massage was performed, resulting in increased cardiac vagal tone and relaxation.[39]

Homeopathy

In a study of 100 cancer patients, individualized homeopathic remedies were prescribed by a practitioner, resulting in reduction of fatigue but not pain.[40] In a randomized, double-blind, placebo-controlled trial of 60 patients with persistent mild traumatic brain injury, each patient was assigned to either individualized homeopathic treatment or placebo, with significant improvement seen for the CAM compared to the control group.[11] Individualized homeopathy was examined in a randomized, placebo-controlled study of otitis media in children. Parental daily symptom diaries demonstrated a significant decrease in symptoms in the homeopathy group.[12] Individualized homeopathic medicine has also been shown to be effective in treating acute childhood diarrhea.[13] In a placebo-controlled, randomized, double-blind study of seborrheic dermatitis in 41 patients, a standardized homeopathic treatment was found to be more effective than a placebo after ten weeks of treatment.[41]

Herbal (Botanicals) Therapy

In a double-blind study, 37 pain patients were randomized to either standardized Arnica treatment or placebo, with a significant reduction in pain seen in the Arnica group relative to control.[42] In a multi-center, double-blind, randomized, placebo-controlled trial, a standardized treatment, *Ruscus aculeatus*, was compared to placebo in 166 women with chronic lower leg venous insufficiency, resulting in significant effects of *Ruscus* extract.[43] A double-blind, placebo-controlled, randomized cross-over clinical trial of 82 patients with sickle cell disease compared standardized NIPRISAN to placebo and found that NIPRISAN significantly reduced the frequency of painful vaso-occlusive crises over a 12-month period.[44] A standardized treatment trial compared the efficacy of the herbal plant, *Orthosiphon grandiflorus* (OG), and the drug sodium potassium citrate (SPC) in 48 patients with renal calculi who were randomized into one of the two arms. After treatment, 90% of the initial clinical symptoms were relieved in both groups, and OG therapy was deemed to be at least as efficacious as SPC in treating renal calculi.[45]

CONCLUSION

To preserve both external and internal validity, the optimal design of CAM interventions in clinical trials appears to be an integrative approach that standardizes an intervention with some room for individualization within the general CAM treatment. One example is a yoga sequence of poses for neck pain with individualization within that series based on individual patient experience, strength, and flexibility. Another example of an integrative approach in Reiki energy therapy might include standardization of the techniques to be used (e.g., magnetic unruffling, correcting chakra spin, specific hand positions) but allowance of flexibility based on individual patient needs (e.g., duration of time for each hand position). Using the example of TCM, individualized treatments are possible, but the replicability of these studies is enhanced if they include a TCM diagnosis, a narrow range of symptoms and diagnoses (e.g., osteoarthropathy-related pain), and an agreed-upon treatment approach based on both Western and Chinese medicine guidelines, with evaluation based on both biomedical and TCM outcomes. Studies that integrate individualization within a standardized framework, with details of the individual treatments recorded, offer the best of both approaches.

The future directions of CAM research will necessitate a shared learning experience of CAM clinicians, biomedical researchers, and likely also social scientists. Cross-fertilization of experiences and knowledge from each realm and paradigm will serve to enhance the scientific rigor of research while still preserving the philosophy of the CAM practice and the utility of the findings for clinical practice. Research training programs and fellowships need to be established for CAM clinicians in academic settings, while investigators need to learn first-hand about the CAM practices that they wish to study. Studies can be designed that allow for both external and internal validity as clinicians and investigators work together toward this goal.

REFERENCES

1. Eisenberg DM, Kessler RC, Foster C, et al. Unconventional medicine in the United States. N Engl J Med 328:246-252,1993.

2. Sharma U. Complementary medicine today: practitioners and patients. London: Tavistock/Routledge, 1992.

3. Merrell WC, Shalts E. Homeopathy. Med Clin North Am 86(1):47-62, 2002.

4. Boissel JP, Ernst E, Fisher P, et al. Overview of data from the homeopathic intervention over no treatment or placebo. In: Report of Homeopathic Medicine Research Group. Brussels: European Commission, 1996.

5. Anthony HM. Clinical research: questions to ask and the benefits of asking them. In: Lewith GT, Aldridge D (eds). Clinical Research Methodology for Complementary Therapies. London: Hodder & Stoughton, 1993.

6. Mercer G, Long AF, Smith U. Researching and evaluating complementary therapies: the states of the debate. Leeds, England, Collaborating Center for Health Service Research, Nuffield Institute for Health, 1995.

7. Nahin RL, Straus SE. Research into complementary and alternative medicine: problems and potential. BMJ 322(7279):161-164, 2001.

8. Critchley JA, Zhang Y, Suthisisang CC, et al. Alternative therapies and medical science: designing clinical trials of alternative/complementary medicines—is evidence-based traditional Chinese medicine attainable? J Clin Pharmacol 40(5):462-467, 2000.

9.Bellavite P, Signorini A. Homeopathy: A Frontier in Medical Science. Berkeley: North Atlantic Books, 1995.

10. Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homoeopathy. BMJ 302(6772):316-323, 1991.

11. Chapman EH, Weintraub RJ, Milburn MA, et al. Homeopathic treatment of mild traumatic brain injury: A randomized, double-blind, placebo-controlled clinical trial. J Head Trauma Rehabil 14:521-542, 1999.

12. Jacobs J, Springer DA, Crothers D. Homeopathic treatment of acute otitis media in children: A preliminary randomized placebo-controlled trial. Pediatr Infect Dis J 20:177-183, 2001.

13. Jacobs J, Jimenes LM, Gloyd SS, et al. Treatment of acute childhood diarrhea with homeopathic medicine: A randomized clinical trial in Nicaragua. Pediatrics 93:719-725, 1994.

14. Richardson J. The use of randomized control trials in complementary therapies: exploring the issues. J Adv Nurs 32(2):398-406, 2000.

15. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 276(8):637-639, 1996.

16. Reilly DT, Taylor MA, McSharry C, et al. Is homoeopathy a placebo response? Controlled trial of homoeopathic potency, with pollen in hayfever as model. Lancet 2(8512):881-886, 1986.

17. Kacperek L. Patients' views on the factors which influence the use of an aromatherapy massage outpatient service. Complementary Therapies in Nursing and Midwifery 3:51-57, 1997.

18. Weiser M, Strosser W, Klein P. Homeopathic vs. conventional treatment of vertigo: A randomized double-blind controlled clinical study. Arch Otolaryngol Head Neck Surg 124:879-885, 1998.

19. Linde K, Jonas WB, Melchart D, et al. The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture. Int J Epidemio1 30(3):526-531, 2001.

20. Jadad AR, Moore RA, Carrol D et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clin Trials 17:1-12, 1996.

21. Linde K, Clausius N, Ramirez G, et al. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. Lancet 350(9081):834-843, 1997.

22. Hsieh, K.H. Evaluation of efficacy of traditional Chinese medicines in the treatment of childhood bronchial asthma: clinical trial, immunological tests and animal study. Pediatr Allergy Immunol 7:130-140, 1996.

23. Fink MG, Wipperman B, Gehrke A. Non-specific effects of traditional Chinese acupuncture in osteoarthritis of the hip. Complement Ther Med 9(2):82-89, 2001.

24. Bensoussan A, Talley NJ, Ring M, et al. Treatment of irritable bowel syndrome with Chinese herbal medicine: a randomized controlled trial. JAMA 280(18):1585-1589, 1998.

25. Lao L, Ezzo J, Bennan BM, et al. In: Stux G, Hammerschlag R (eds). Clinical Acupuncture: Scientific Basis. New York: Springer-Verlag, 2001, pp. 187-209.

26. Garfinkel MS, Singhal A, Katz W A, et al. Yoga-based intervention for carpal tunnel syndrome: a randomized trial. JAMA 280(18):1601-1603, 1998.

27. Singh V, Wisniewski A, Britton J, et al. Effect of yoga breathing exercises (pranayama) on airway reactivity in subjects with asthma. Lancet 335(8702):1381-1383, 1990.

28. Manchanda SC, Narang R, Reddy KS, et al. Retardation of coronary atherosclerosis with yoga lifestyle intervention. J Assoc Physicians India 48:687-694, 2000.

29. Meehan TC. Therapeutic touch and postoperative pain: a Rogerian research study. Nurs Sci Q 6(2):69-78, 1993.

30. Turner JG, Clark AJ, Gauthier DK, et al. The effect of therapeutic touch on pain and anxiety in burn patients. J Adv Nurs 28(1):10-20, 1998.

31. Zeltzer L, LeBaron S. Hypnosis and nonhypnotic techniques for reduction of pain and anxiety during painful procedures in children and adolescents with cancer. J Pediatrics 101(6):1032-1035, 1982.

32. Stradling J, Roberts D, Wilson A, et al. Controlled trial of hypnotherapy for weight loss in patients with obstructive sleep apnea. Int J Obes Relat Metab Disord 22(3):278-281, 1998.

33. Ewer TC, Stewart DE. Improvement in bronchial hyper-responsiveness in patients with moderate asthma after treatment with a hypnotic technique: a randomised controlled trial. Br Med J (Clin Res Ed) 293(6555):1129-1132, 1986.

34. Schubert DK. Comparison of hypnotherapy with systematic relaxation in the treatment of cigarette habituation. J Clin Psychol 39(2):198-202, 1983.

35. Zeltzer L, LeBaron S, Zeltzer PM. The effectiveness of behavioral intervention for reduction of nausea and vomiting in children and adolescents receiving chemotherapy. J Clin Oncol 2(6):683-690, 1984.

36. Grealish L, Lomansney A, Whiteman B. Foot massage: a nursing intervention to modify the distressing symptoms of pain and nausea in patients hospitalized with cancer. Cancer Nursing 23(3):237-243, 2000.

37. Okvat HA. Oz MC, Ting W, et al. Massage therapy for patients undergoing cardiac catheterization. Altern Ther Health Med 8(3):68-75, 2002.

38. Le Blanc-Louvry I, Costaglioli B, Boulon C, et al. Does mechanical massage of the abdominal wall after colectomy reduce postoperative pain and shorten the duration of ileus? Results of a randomized study. J Gastrointest Surg 6(1):43-49, 2002.

39. Delaney JP, Leong KS, Watkins A, et al. The short-term effects of myofascial trigger point massage therapy on cardiac autonomic tone in healthy subjects. J Adv Nurs 37(4):364-371, 2002.

40. Thompson EA, Reilly D. The homeopathic approach to symptom control in the cancer patient: a prospective observational study. Palliat Med 16(3):227-233, 2002.

41. Smith SA, Baker AE, Williams JH. Effective treatment of seborrheic dermatitis using a low dose, oral homeopathic medication consisting of potassium bromide, sodium bromide, nickel sulfate, and sodium chloride in a double-blind, placebo-controlled study. Altern Med Rev 7(1):59-67, 2002.

42. Jeffrey SL, Belcher HJ. Use of Arnica to relieve pain after carpal-tunnel release surgery. Altern Ther Health Med 8(2):66-68, 2002.

43. Vanscheidt W, Jost V, Wolna P, et al. Efficacy and safety of a Butcher's broom preparation (Ruscus aculeatus L. extract) compared to placebo in patients suffering from chronic venous insufficiency. Arzneimittelforschung 52(4):243-250, 2002.

44. Wambebe C, Khamofu H, Momoh JA, et al. Double-blind, placebo-controlled, randomised cross-over clinical trial of NIPRISAN in patients with Sickle Cell Disorder. Phytomedicine 8(4):252-261, 2001.

45. Premgamone A, Sriboonlue P, Disatapomjaroen W, et al. Long-term study on the efficacy of a herbal plant, Orthosiphon grandiflorus, and sodium potassium citrate in renal calculi treatment. Southeast Asian J Trop Med Public Health 32(3):654-660, 2001.

Acknowledgements:

We appreciate the helpful comments of the following CAM practitioners: Michael Waterhouse, MA, LAc; Evan Ross, LAc, DaM; Ka-Kit Hui, MD; Cynthia D Myers, PhD; Claire Cassidy, PhD; Marion Wolfe Dixon, MA, LMT; Ray Siderius, LAc; Beth Sternlieb; Francesca De Picciotto, MD; Audrey Easton, PhD; Coleen Werner; Lobsang Rapgay, PhD; and Katherine DePlanques, PhD.

ASSESSMENT OF CANCER-RELATED SYMPTOMS: RELEVANCE FOR CAM RESEARCH METHODOLOGY

Charles S. Cleeland, PhD*
McCullough Professor of Cancer Research
Director, Pain Research Group
University of Texas
M.D. Anderson Cancer Institute

Karen O. Anderson, PhD
Assistant Professor, Pain Research Group
M.D. Anderson Cancer Center

Pain Research Group, Division of Anesthesiology and Critical Care, University of Texas M. D. Anderson Cancer Center, Houston, TX

Supported by Public Health Service grants CA64766 and CA26582 from the National Cancer Institute, DAMD17-94-J-4233 from the U.S. Army Medical Research and Materiel Command, and SIG #21 from the American Cancer Society

**Corresponding author:*
Charles S. Cleeland, PhD
Director, Pain Research Group
University of Texas
M.D. Anderson Cancer Institute
1100 Holcombe Blvd, Box 221
Houston, TX 77030

ASSESSMENT OF CANCER-RELATED SYMPTOMS: RELEVANCE FOR CAM RESEARCH METHODOLOGY

INTRODUCTION

Patients with cancer typically experience multiple symptoms related to cancer and cancer treatment. These symptoms can include physical (e.g., nausea, dyspnea), cognitive (e.g., memory problems, impaired concentration), and affective (e.g., sadness, anxiety) experiences associated with the disease and its treatments. Symptom severity is related to the extent of disease and the aggressiveness of therapies such as surgery, chemotherapy, radiotherapy, and biological therapies. Common symptoms of cancer and cancer treatment significantly impair the daily function and quality of life of patients. Pain is a good example: When pain is present, it adversely affects patients' mood, activity, and ability to relate to others (Serlin et al., 1995). Similarly, fatigue, nausea and vomiting, shortness of breath, and psychological distress can add tremendously to the burden of having a life-threatening disease (Portenoy et al., 1994, Cleeland, et al., 2000).

Unrecognized symptoms may become so severe that an emergency room visit and even hospitalization are required for management, adding substantially to the cost of treatment and to the disruption of the patient's routine and that of the family. Untreated symptoms may also interrupt cancer treatments and negatively influence disease-related outcomes. Moreover, symptoms often have an adverse effect on patients' satisfaction with their medical care. In a recent survey of over 900 cancer patients, 73% reported satisfaction with their cancer treatment, but only 46% were satisfied with their symptom management (Ashbury et al., 1998).

Relief of symptoms is one of the most frequent reasons that patients turn to complementary and alternative medical treatment (CAM). Fatigue is one symptom that often motivates patients to seek help with CAM. A recent study of HIV-infected patients found that the two predictors of supplement purchase were higher education and greater fatigue (Fairfield et al., 1998). Similarly, cancer patients seek relief from fatigue, pain, disturbed sleep, depression, and other symptoms using supplements and a variety of behaviorally based therapies.

Despite the tremendous distress that symptoms can cause, symptom assessment is rarely a part of routine cancer care. Physicians and nurses are often hard-pressed to find time to assess symptoms in a hectic clinic schedule. Healthcare professionals may wait until patients spontaneously complain of symptoms before formally assessing them. Some may be unfamiliar with assessment that depends on subjective report. Moreover, physicians and nurses are more apt to rely on their own experience or professional judgment instead of using a standardized assessment instrument. The lack of systematic assessment and treatment of symptoms is expected to be a major reason that patients seek help from CAM.

Both the wide-spread current use of CAM for symptom management, and the promise of many CAM-based therapies to relieve symptom distress underlines the necessity of well-designed clinical trials to determine which of these methods might be effective in the control of symptoms. Such trials are obviously critically dependent on reliable, valid, and practical measures of symptom severity and impact. In this chapter, we provide an overview of symptom assessment in

oncology and discuss the characteristics of a good symptom assessment tool. We also briefly review the available multiple symptom assessment measures that can be used to document change or lack of change that might be attributed to a treatment. Innovative trends in symptom measurement are presented.

PREVALENCE OF CANCER-RELATED SYMPTOMS

As background, we will summarize studies of the prevalence, severity, and impact of symptoms related to cancer and its treatment. A number of studies have examined the symptom patterns of cancer patients who have advanced disease. The most frequently examined symptom has been pain. A study completed in the Eastern Cooperative Oncology Group surveyed over 1,300 patients with recurrent or metastatic cancer (Cleeland et al., 1994). Sixty-seven percent of the patients had pain or were being treated for pain with daily analgesics. Thirty-six percent of the total sample had pain that was reported as substantially interfering with their daily Jives. Of those patients with pain, 42% were receiving analgesics that were inadequate for the intensity of their pain. Other studies of metastatic cancer patients have found that large numbers of patients experience significant pain that requires treatment until death (Brescia et al., 1990; Coyle et al., 1990; Fainsinger et al., 1991).

Several studies have examined other cancer-related symptoms in patients with advanced disease. Schuit and colleagues found that pain, constipation, nausea, anorexia, coughing, and dyspnea were frequent symptoms in outpatients with advanced cancer (Schuit et al., 1998). A retrospective study of over 1,000 advanced-stage cancer patients admitted to a palliative care hospital found that anorexia, cognitive impairment, nausea, dyspnea, and dysphagia, as well as pain, were reported by significant percentages of patients (Brescia et al., 1990). The percentages of patients reporting these symptoms ranged from 19% for nausea to 38% for pain. Coyle et al. (1990) found that fatigue, weakness, pain, sleepiness, and cognitive impairment were frequent symptoms in their study of terminal patients enrolled in a supportive care program. Fatigue (58%) and pain (54%) were the most prevalent symptoms. Similarly, a prospective study of cancer patients in palliative care centers in Europe, Australia, and the United States found that pain (57%) and weakness (51%) were the most frequently reported symptoms (Vainio et al., 1996). Weight loss, anorexia, constipation, nausea, and dyspnea also were common.

Very few epidemiological studies have examined multiple symptoms of cancer patients with early stage disease. However, symptoms associated with aggressive treatments such as chemotherapy and radiotherapy have been reported. For example, multiple studies have found that the majority of patients undergoing chemotherapy or radiotherapy report significant fatigue during the course of treatment (Hickok et al., 1996; Irvine et al., 1994; Jacobsen et al., 1999; Smets et al., 1996). A few studies have assessed multiple symptoms in samples of cancer patients with a variety of disease stages. Portenoy and colleagues administered the Memorial Symptom Assessment Scale to a random sample of inpatients and outpatients with breast, prostate, colon, or ovarian cancer (Portenoy et al., 1994). Thirty-seven percent of the patients had local disease or no current evidence of active cancer. The most frequently reported symptoms for the entire sample were lack of energy, worry, feeling sad, pain, feeling nervous, drowsiness, and dry mouth. The patients with metastatic disease reported more symptom distress than patients with

less advanced disease. Fatigue also was the most prevalent symptom in a sample of newly diagnosed cancer patients, 29% of whom had early stage disease (Degner et al., 1995).

The Pain Research Group recently administered the M. D. Anderson Symptom Inventory to a sample of over 500 outpatients with diverse cancer diagnoses and disease stages (Cleeland et al., 2000). Twenty-nine percent of the patients had metastatic disease. The symptoms of highest mean intensity were fatigue, not getting things done, weakness, worry, emotional distress, sleep disturbance, drowsiness, and lack of appetite.

Quality-of-life studies of cancer survivors have found that many patients continue to experience physical, affective, or cognitive symptoms even when their disease is in remission and treatment has ended. These symptoms may be due to physiological changes associated with prior treatments, delayed side effects of treatment, or long-term consequences of the disease. For example, many survivors of bone marrow transplantation report cognitive impairment, physical symptoms, or emotional distress many years after the transplant (Andrykowski et al., 1995; McQuellon et al., 1996; Prieto et al., 1996).

WHY MEASURE SYMPTOMS?

Good management of cancer-related symptoms is based on regular and accurate symptom assessment and effective communication between patients and physicians. Problems with assessment are recognized by those who treat cancer patients as a major barrier to good symptom control. Medical oncologists were surveyed in a study conducted by the Eastern Cooperative Oncology Group (Von Roen et al., 1993). Seventy-five percent of the physicians indicated that the most important barrier to cancer pain management was inadequate pain assessment. Over 60% of physicians cited the unwillingness of patients to report pain as a major barrier. A similar study was completed with physician members of the Radiation Therapy Oncology Group (Cleeland et al., in press). As with the ECOG physicians, members of the radiation oncology group perceived poor pain assessment to be the major barrier to pain management (77%), followed by patient reluctance to report pain (60%).

Several studies of outpatients with metastatic or recurrent cancer receiving treatment at ECOG institutions found that large percentages of the patients with pain were receiving inadequate analgesic treatment (Cleeland et al., 1994, 1997). A discrepancy between the physician and patient in rating the severity of the patient's pain was a major predictor of under-medication for pain (Cleeland et al., 1994). In addition, patients seen at centers that predominantly treated minority patients were three times more likely to have inadequate pain management than those treated elsewhere. Other factors that predicted inadequate pain treatment included age of 70 years or older, female sex, and better performance status. These results suggest that careful and accurate symptom assessment is particularly important for cancer patients from minority groups, elderly patients, female patients, and patients who appear to be functioning well.

The results of a recent study of cancer outpatients demonstrated that standardized pain assessment can lead to improved pain management. The patients were asked to rate their average and worst pain during the previous week, as well as satisfaction with their pain management and the degree of pain relief received (Trowbridge et al., 1997). The oncologists treating the patients in the intervention group were asked to read a summary of the patients' pain-scale scores prior to

seeing the patients. The summary was not available for the oncologists treating the control group patients. The results revealed a decrease in the incidence of pain in the intervention group. In addition, the oncologists treating the intervention group were more likely to make changes in the patients' analgesic prescriptions, as compared to the oncologists treating the control group patients.

GENERAL GUIDELINES FOR SYMPTOM ASSESSMENT

Patients should be asked to describe their symptoms in their own words. If a patient has difficulty communicating, then a family member can be asked to provide additional information on the history of the symptom. Any previous pharmacological or nonpharmacological treatment for a symptom should be determined, as well the patient's response to the treatment. It is helpful to understand the temporal pattern of the symptom, when it occurs, and whether it is constant or episodic. Factors that aggravate or alleviate a symptom should be determined. As symptoms may occur in clusters, it should be determined if two or more symptoms tend to occur together. Additional laboratory or other diagnostic tests may be necessary in order to understand the etiology and pathophysiology of specific symptoms (Gonzales et al., 1991).

Although the initial medical evaluation will often reveal the presence of patients' symptoms, patients may be afraid to report symptoms verbally to a health care provider, for a variety of reasons. They may feel that reporting symptoms is a type of complaint or a criticism of the physician's or nurse's clinical skills. Thus, complaining of symptoms may defeat their attempts to be a "good" patient. Differences in the ethnic and cultural backgrounds of patients and health care providers may hinder effective communication about symptoms. In addition, both patients and health care professionals recognize that symptoms are themselves a negative prognostic indicator (Chang et al., 1998). Patients may feel that reporting symptoms distracts the physician's attention from taking care of the cancer or is an admission that their disease is growing worse (Ward et al., 1993). Patients may also worry that discussing a symptom will lead to a new set of medications with unknown side effects.

The use of standardized assessment instruments encourages patients to report their symptom's in a non-threatening format. The use of simple measurement scales greatly improves the symptom assessment process, helps direct treatment choices, and provides information about the effectiveness of treatment. While some symptoms such as fatigue are difficult to treat, other symptoms such as pain can be treated in a relatively straightforward manner. Evidence-based guidelines and clinical pathways are available to guide pain treatment (Jacox et al., 1994; American Pain Society, 1999). Most treatment guidelines for the management of cancer pain assume that pain severity scales will be used to assess the patient's pain experience. The severity of the patient's pain helps to determine the appropriate treatment.

The relationship between the intensity of a symptom and the degree of the symptom's interference in functions such as activity and mood can help to determine the level of symptom severity. For example, a recent study found that mild, moderate, and severe pain differentially affect the function of cancer patients (Serlin et al., 1995). Based on the degree of interference with function, pain intensity ratings of 1–4 on a 0–10 scale correspond to mild pain, 5–6 to moderate pain, and 7–10 to severe pain. A similar methodology can be applied to other cancer-related symptoms to determine levels of severity that may impact treatment decisions.

A recent study of fatigue in cancer patients found that ratings of fatigue intensity of 7 to 10 on a 10-point scale corresponded to severe fatigue, as determined by interference with function (Mendoza et al., 1999). Additional research is needed to specify the boundary between mild and moderate fatigue. These studies suggest that, at least in terms of interference, symptom severity scales are non-linear, and that an objective of a symptom intervention might be to reduce symptom severity to the mild range instead of eliminating the symptom altogether. At least for cancer patients, completely eliminating a symptom might be impossible, or only possible by also producing undesirable side effects (such as sedation) due to the symptom intervention.

Single-symptom scales, such as those measuring fatigue or nausea, can be useful in documenting the severity of a symptom and in guiding treatment recommendations (Mendoza et al., 1999; Morrow, 1988). However, most cancer patients with advanced disease experience multiple symptoms. Multiple-symptom scales can be used to monitor patients during and after active treatment and to identify symptoms that require interventions. Several multiple symptom scales have been developed, and their usefulness has been demonstrated in epidemiologic studies (Portenoy et al., 1994a, Portenoy et al., 1994b), in predicting survival (Chang et al., 1998), and in aiding clinical care (Jenkins et al., 1998).

CHARACTERISTICS OF A GOOD SYMPTOM ASSESSMENT MEASURE

Any acceptable symptom measure must demonstrate adequate reliability and validity. Reliability refers to the consistency or reproducibility of measurement. When patients are asked to rate symptoms repeatedly over short intervals (test-retest reliability), the correlations between the ratings should be high. Of course the symptoms of cancer patients can change during short intervals; so test-retest reliability is best measured for patients whose symptoms and disease status are relatively stable. Another measure of reliability is internal consistency or the degree to which individual items in a measure correlate with the total score. Cronbach's alpha is one measure of internal consistency that is often reported for assessment measures.

Validity refers to whether the assessment instrument is actually measuring what it is designed to measure. There are several types of validity that are important to consider in symptom assessment. Content validity assesses whether the instrument adequately covers the area of interest. For example, symptom measures designed to measure multiple cancer-related symptoms should cover the range of cancer symptoms and include the most prevalent and clinically important symptoms. Content validity is related to face validity, the judgment of the relevant professional community that the instrument really measures what it is intended to measure.

An assessment instrument has construct validity if it appears to measure the abstract construct or idea of interest. Construct validity is often established using hypothesis testing. For example, it has been hypothesized and demonstrated that patients with advanced disease will report more symptoms and more distress due to symptoms than patients with early stage disease (Cleeland et al., 1999; Portenoy et al., 1994). The results of such studies provide evidence of the construct validity of the symptom assessment measure used in the research. Criterion-related validity is determined by correlating the assessment measure with a known "gold standard" for assessing

the variable of interest, the criterion. Unfortunately, there is no gold standard available for measuring symptoms.

Symptoms are subjective experiences of the patient, and we are dependent on patients to report the characteristics of their symptoms. Most symptom assessment measures are self-administered questionnaires that patients can complete in a clinic, hospital, or home setting. Staff members should be available in person or by telephone to answer patients' questions if they have difficulty completing the questionnaire. Administration of the symptom measure in an interview format by a staff member may be necessary for some patients. Telephone administration is another option if patients are not able to come into the clinic setting.

The ideal symptom assessment tool for cancer patients should include symptoms related to cancer and its treatment that occur frequently and are distressing for patients. A good symptom measure needs to include symptoms that are important to monitor in a clinical setting and that patients may be reluctant to report (e.g., pain, depression). The ideal symptom assessment measure also should be short in order to place minimal burden on the sick patients who are asked to complete it. Brevity is particularly important if the symptom scale is used on a frequent basis to monitor symptoms. A symptom measure must be easy to understand so that patients with limited educational backgrounds can complete it with minimal assistance. Availability in multiple languages is important, especially in settings where patients come from diverse racial and ethnic groups. In addition, a good symptom assessment measure should be applicable in both clinical and research settings.

Symptom assessment involves asking the patient to rate a given symptom using some type of response scale. The simplest type of scale is a dichotomous one that asks the patient to indicate the presence or absence of a particular symptom. This type of response, however, does not provide any information about the characteristics of a symptom. Most response scales included in symptom measures ask patients to rate a symptom on a graded scale that provides information about intensity, frequency, or distress. The severity (i.e., intensity) of symptoms can be rated using Visual Analogue Scales (VAS), Verbal Rating Scales (VRS), and Numeric Rating Scales (NRS).

The VAS usually consists of a 10-cm horizontal straight line. One end of the line represents "no symptom" and the other end represents some concept such as "symptom as bad as you can imagine." The patient is asked to place a mark across the line indicating how much of the scale is equivalent to the severity (or frequency, unpleasantness, etc.) Visual analog scales have been used extensively in pharmacological and other types of symptom research. Considerable evidence of their reliability and validity has been found. However, the VAS requires that patients understand the analog concept. These scales often necessitate considerable instruction and supervision from clinical or research staff. It has been suggested that VAS are not appropriate for patients with limited educational backgrounds (Cleeland, 1991). The VAS also requires additional time for scoring of the patient's responses.

Verbal rating scales are category scales that ask the patient to choose a verbal descriptor which best describes a symptom. A simple VRS consists of "none," "mild," "moderate," and "severe." Although these scales can be useful in symptom assessment, the VRS assumes that patients have

approximately the same meaning in mind when choosing the descriptor that best describes their symptom. This assumption is often questionable given that patients come from very diverse educational, cultural, and linguistic backgrounds.

The numeric rating scales (NRS) ask patients to rate their symptoms on simple numerical scales. The scales typically range from 0 to 5, 0 to 10, or 0 to 100. Numeric scales usually are presented horizontally, with a verbal anchor given for each end of the scale. For example, the "0" end of a symptom scale assessing pain may be defined as "no pain," and the other end may be defined as "pain as bad as you can imagine" or "very severe pain." The NRS are usually easier for patients to understand than VAS. The use of numbers instead of words may remove some sources of cultural and linguistic variation (Cleeland, 1990, 1997).

The NRS, VAS, and VRS have been employed extensively in pain research. In pain assessment in clinical settings, these three scales of severity provide nearly equivalent data (Jensen, Karoly, & Braver, 1986). Given this equivalency of information, clarity, ease of administration, and simplicity of scoring become justified criteria in selecting a response scale. All three scales are highly intercorrelated, although the NRS and VAS are most highly correlated with one another (Syrjala, 1987).

In clinical research, the NRS has been found to be more reliable than the VAS, especially for patients with limited education (Ferraz, et al., 1990). For very ill patients, oral versions of the NRS can be administered. Numerical scales have been endorsed for use in cancer clinical trial instruments because they are easier to understand and score than VAS or VRS (Moinpour et al., 1989). Eleven-point rating scales (0–10) maximize a trade-off between subject ease of responding and increasing reliability with longer numerical scales (Nunnally, 1978).

Impact of Cancer-Related Symptoms

Symptoms have a significant impact on patients' function and quality of life. A recent study of Chinese cancer patients found that pain had an adverse impact on quality of life, even after controlling for disease severity (Wang et al., 1999). The assessment of patients' symptoms should include some evaluation of the impact of the symptoms on patients' daily lives and general well-being. The assessment of symptom impact on function can be determined in several ways. The M. D. Anderson Symptom Inventory includes items that assess the amount of interference due to symptoms in several domains: general activity, walking, work, sleep, mood, relations with others, and enjoyment of life. A symptom interference score is easily calculated from these items (Cleeland et al., 2000). If additional evaluation of physical function is needed, assessment instruments that focus on physical function, such as the Sickness Impact Profile, are available (Bergner et al., 1981).

Multiple quality of life measures have been developed and are frequently used as outcome measures in clinical trials with cancer patients (Cella, 1996; Moinpour, 1994; Spilker, 1996). A review of these measures is beyond the scope of this chapter. Quality of life measures typically include the assessment of physical, emotional, and social functioning. A few symptom items may be included, but these measures do not screen effectively for cancer-related symptoms. Such assessment tools can be used to evaluate general health-related quality of life and to identify

areas that may require additional assessment. However, many quality of life instruments are lengthy and place an additional burden on cancer patients. Symptom interference scores are highly related to scores on quality of life scales, and conceptually might be thought of as accounting for the variance of symptoms in patients' reports of health-related quality of life.

MULTI-SYMPTOM SCALES FOR ASSESSMENT OF CANCER-RELATED SYMPTOMS

Symptom Distress Scale (SDS)

McCorkle and colleagues developed one of the first measures for the assessment of multiple symptoms, the Symptom Distress Scale or SDS (McCorkle & Young, 1978). Symptom distress is defined as the degree of discomfort reported by patients in relation to their perceptions of symptoms. The original SDS included ten symptoms: nausea, mood, appetite, insomnia, pain, mobility, fatigue, bowel pattern, concentration, and appearance. Each symptom is rated on a 5-point Likert scale with verbal descriptors anchoring each end of the scale. A score of 5 represents a maximum amount of distress for a particular symptom, and a score of 1 represents no distress for a given symptom. For example, a score of 5 on the nausea scale is described as "I feel as sick as I could possibly be" and a score of 1 is described as "I do not feel sick at all." Patients are asked to circle the number that most closely represents how they feel at that moment or on that particular day. A total symptom distress score can be obtained by adding the scores for all ten symptoms. In the initial validation study, over 60% of the cancer patients who completed the scale reported a high level of distress (score of 4 or 5) for at least one of the symptoms.

The SDS was subsequently modified to include more of the symptoms typically experienced by patients with lung cancer and other chronic illnesses (McCorkle, 1983). The 13 items on this modified Symptom Distress Scale include: nausea (frequency), nausea (intensity), appetite, insomnia, pain (frequency), pain (intensity), fatigue, bowel patterns, concentration, appearance, breathing, outlook, and cough. Verbal descriptors were added for each point on the 5-point Likert scales. When the modified SDS was given to lung cancer and myocardial infarction patients, cancer patients reported more symptom distress than the cardiac patients (McCorkle, 1983).

The modified SDS demonstrates a number of strengths. The scale is short enough for sick patients to complete in less than 10 minutes. Research results have demonstrated that the scale has good reliability and validity. Scores on the scale correlated with general measures of health status and demonstrated sensitivity to change following treatment (McCorkle, 1987). The total symptom distress score on the SDS was a significant predictor of survival in lung cancer patients (Degner, Sloan, 1995). However, there are problems associated with the SDS. It is not clear whether all of the symptom items measure distress due to symptoms. Most of the symptom anchors refer to frequency or intensity of symptoms. Moreover, Likert scales with verbal descriptors may be confusing for patients. Finally, the SDS contains only one affective symptom (outlook) and may not adequately screen for emotional distress.

A modified version of the SDS, the Symptom Experience Scale, was developed to assess the symptoms associated with treatment of breast cancer (Samarel et al., 1996). The 24-item

Symptom Experience Scale (SES) includes 8 symptoms from the SDS and asks patients to rate separately the frequency, intensity, and distress associated with each symptom. The SES takes longer to complete than the SDS and does not include affective symptoms such as sadness and anxiety that are often reported by individuals with cancer.

The Memorial Symptom Assessment Scale (MSAS)

The MSAS evaluates 32 physical and psychological symptoms, asking patients to rate separately the frequency, intensity, and distress of each symptom during the past week (Portenoy et al., 1994). Symptom frequency is rated on a 4-point Likert scale from "rarely" to "almost constantly," and symptom severity is rated on a 4-point Likert scale from "slight" to "very severe." Symptom distress is rated on a 5-point Likert scale from "not at all" to "very much." A summary index, the MSAS Global Distress Index, is the mean of 10 frequently endorsed items. Subscale scores that measure physical (MSAS-Phys) and psychological (MSAS-Psych) symptom distress also can be calculated.

In the initial validation study, the MSAS was administered to over 200 inpatients and outpatients with prostate, breast, colon, or ovarian cancer. A factor analysis yielded two factors that distinguished three major symptom groups and several subgroups. The major groups comprised psychological symptoms (e.g., "worrying," "feeling sad"), high prevalence physical symptoms (e.g., "lack of energy," "pain"), and low prevalence physical symptoms (e.g., "numbness/tingling in hands/feet," "cough"). The Global Distress Index and the physical and psychological subscale scores were significantly correlated with measures of clinical status and quality of life. The physical symptom subscale score on the MSAS has also been shown to be predictive of survival in a sample of lung cancer patients (Chang et al., 1998).

The MSAS is a multidimensional tool that provides a comprehensive assessment of the prevalence and characteristics of a wide spectrum of symptoms. It is a valuable research tool that has been used with patients with various types of cancers (Harrison et al., 1997; Kornblith et al., 1995). It has been suggested that the MSAS can be used as an outcome measure in cancer clinical trials (Ingham & Portenoy, 1996). A short form of the MSAS (MSAS-SF) that includes intensity or frequency ratings of 32 symptoms has been developed (Chang et al., 1997). However, even the short form may be too long for very ill patients to complete, and it is impractical for the rapid assessment of symptoms in a clinical setting.

The Rotterdam Symptom Checklist (RSC)

This 30-item scale was designed to measure the symptoms of cancer patients who participate in clinical trials (de Haes, van Knippenberg & Neijt, 1990). Patients are asked to rate on a 4-point Likert scale the extent to which a symptom bothered them during the past week. Examples of symptoms on the checklist are: lack of appetite, irritability, tiredness, depressed mood, and nausea. The RSC also contains 8 items assessing activities of daily living. The validation sample included cancer patients receiving treatment, disease-free patients, and normal controls. Factor analysis yielded two dimensions, a psychological and a physical dimension. Reliability as measured by Cronbach's alpha was high for both factors.

An advantage of the RSC is that it can be adapted for use with various patient groups by adding or deleting specific items. The major disadvantages of this tool include its length and its use of a rating system based on verbal descriptors. Scales based on verbal descriptors may be difficult for some patients to understand and may not be interpreted in the same way by all patients. Another difficulty with the RSC relates to the measurement of pain. The scale contains items assessing pain in specific body areas but does not include pain as a general symptom.

The Edmonton Symptom Assessment System (ESAS)

The ESAS was designed for the regular assessment of patients in a palliative care unit (Bruera, 1991). This inventory consists of nine visual analogue scales (0–100 mm) that measure the patient's current level of pain, activity, nausea, depression, anxiety, drowsiness, appetite, sensation of well-being, and shortness of breath. Numerical rating scales (0–10) can be substituted for the VAS. A general symptom distress score on the ESAS is obtained by adding the scores of all nine scales together. A tenth visual analog or numerical scale is available for patients to add any other symptom that is troublesome for them. The ESAS can be completed by the patient either with or without assistance, or it can be completed by the patient's nurses or relatives when self-report is not possible. In the initial study, 84% of the patients were able to complete the ESAS at least once during their stay in the palliative care unit. However, shortly before death, 83% of the assessments were completed by a nurse or relative instead of the patient.

The 9-item ESAS was recently modified for use with an Australian population (Philip et al., 1998). The scale to measure activity was replaced with a VAS to measure weakness. In addition, a VAS was added to measure the percentage of pain relief that patients received in the past 24 hours from pain treatments. The modified ESAS proved to be a valid self-administered scale when evaluated in several inpatient and outpatient palliative care settings.

The ESAS is a short and useful tool for the regular assessment of symptoms in palliative care patients. The system has been used successfully to monitor treatment outcomes in palliative care programs. The ESAS has not been validated in outpatient samples of cancer patients who are not receiving primarily palliative care.

M.D. Anderson Symptom Inventory (MDASI)

The Pain Research Group at the University of Texas M. D. Anderson Cancer Center has developed a multiple-symptom assessment measure, the M. D. Anderson Symptom Inventory (Cleeland et al., 2000). We initially generated a list of 26 symptoms derived from existing symptom assessment scales as well as from consultation with a working group of oncology professionals (medical and radiation oncologists and oncology nurses). The item pool was administered to two samples of cancer outpatients and an inpatient sample of blood and marrow transplantation patients. The number of symptoms in the MDASI was reduced using statistical methods and clinical judgment. Symptoms that were redundant with other symptoms (e.g., weakness, redundant with fatigue) or extremely low frequency (e.g., bleeding) were eliminated from the core inventory. Symptoms that were predictive of overall symptom distress and/or were

judged to be important clinical indicators were retained, even if they were highly correlated with other symptoms (e.g., vomiting and nausea).

The goal was to develop a core list of symptoms that each patient rates, plus modules of additional symptoms that can be included for patients who are at risk for symptoms not highly prevalent in oncology patients in general. The core symptom items on the MDASI can be used to monitor patients' symptoms in routine clinical care. Subsets of additional symptom items can be added to the basic MDASI for patients who are receiving aggressive treatments (e.g., bone marrow transplantation) or who have cancer diagnoses associated with specific symptoms (e.g., lung cancer and coughing).

The core MDASI consists of 13 symptoms: pain, fatigue, nausea, sleep disturbance, emotional distress, shortness of breath, lack of appetite, drowsiness, dry mouth, sadness, vomiting, difficulty remembering, and numbness or tingling. Each symptom is rated on an 11-point scale, with 0 being "not present" and 10 being "as bad as you can imagine." The MDASI also contains six items that describe how much the symptoms have interfered with different aspects of the patient's life during the past 24 hours: general activity, mood, walking ability, normal work (including work outside the home and housework), relations with other people, and enjoyment of life. Each interference item is rated on an 11-point scale with 0 being "does not interfere" and 10 being "completely interferes."

The MDASI uses 11-point scales to assess symptoms and interference in daily life due to symptoms for several reasons. First, 0 to 10 ratings of symptom severity are used fairly routinely to assess symptom severity in the clinic. The numeric scales are easily administered and can be comprehended by most patients. Second, they appear to produce equivalent data when presented in different languages. Third, they can be used with computer-assisted telephone assessment of patients at home in what is sometimes called an interactive voice response (IVR) mode. The MDASI asks patients to rate the intensity of each symptom at its worst in the last 24 hours. The MDASI does not include the additional ratings of symptom distress and frequency that are found in longer symptom scales such as the MSAS. The MDASI is designed for simplicity, brevity, and acceptability to very ill patients.

In the initial validation study, the core MDASI demonstrated good reliability and sensitivity to disease severity and treatment status (Cleeland et al., 1999). For example, the mean symptom severity and symptom interference ratings of patients with poor performance status were significantly higher than those of patients with good performance status. Patients who were undergoing active treatments demonstrated expected differences in symptom severity and symptom interference, as compared to patients being seen only for follow-up visits. In sum, the core MDASI is a psychometrically sound, brief assessment measure that can be used to screen patients in the clinic or hospital setting. The disease and treatment-specific modules for the MDASI (e.g., blood and marrow transplantation, lung cancer) are currently being developed and evaluated.

Multiple-symptom scales are often highly desirable as outcome instruments. First, treatments may have effects on multiple symptoms simultaneously, and it is important to be able to document the potential multiple symptom advantages of a treatment. Second, many potential

treatments, including CAMs, are felt to have a general positive effect for the patient, but only by assessing multiple symptom domains can the specific targets of the treatment be identified. For example, depression may but fatigue may not be improved by a given intervention (Murrow et al., 2001). Finally, as we become more knowledgeable about the mechanisms underlying symptom expression, it is likely that several symptoms might be jointly affected by a given intervention. For example, the inflammatory cytokines, such as IL1 and IL6, might simultaneously promote pain, fatigue, depression, and cognitive impairment.

ASSESSMENT OF INDIVIDUAL CANCER-RELATED SYMPTOMS

In some cases, however, clinical researchers may need additional information about specific symptoms. Additional data are particularly helpful when a symptom is difficult to treat and/or very distressing for the patient, or when it is reasonable to assume that a specific intervention might have a special benefit for a specific intervention, such as an analgesic for pain.

Many assessment instruments are available for some individual cancer-related symptoms, such as pain, fatigue, and depression. The assessment of other cancer- related symptoms such as anorexia or drowsiness is more difficult due to a lack of validated assessment tools. When a validated tool is not available, clinicians can use the individual symptom item from a multiple symptom scale for regular assessment and to monitor treatment progress. For example, a numerical rating of drowsiness on a 0–10 scale can be used to monitor this symptom that is a frequent side effect of opioid medications.

A full review of individual symptom assessment scales is beyond the scope of this chapter. We will discuss the instruments that were developed for the assessment of two common cancer-related symptoms: pain and fatigue. Our focus will be on instruments that were developed specifically for cancer symptom assessment or have been validated with cancer patients.

Pain Assessment

Many assessment instruments are available for the measurement of pain, but most of these instruments were not designed specifically for the assessment of cancer- related pain. An exception is the Brief Pain Inventory (BPI), a questionnaire which was specifically designed to assess cancer pain (Cleeland, 1989). The BPI provides information on the intensity of pain as well as the degree to which pain interferes with patients' function. The BPI uses 0 to 10 numeric rating scales. Since pain due to cancer can be quite variable over a day, the BPI asks patients to rate their pain intensity at the time of responding ("pain now"), and also at its worst, least, and average over the previous week. The ratings can also be made for the last 24 hours. The BPI asks patients to rate how much pain interferes with their general activity, walking, normal work, relations with other people, mood, sleep, and enjoyment of life. The mean of the patient's ratings on the pain interference items can be used to determine a pain interference score. As ratings on the "pain worst" item increase, additional pain interference items typically are rated as impaired. The BPI also asks patients to rate the amount of relief that they feel their current pain treatment provides. This item can be considered an indication of patients' satisfaction with their current pain treatment. In addition, the BPI asks patients to provide a graphic representation of the

location of their pain. The BPI includes a front and back view of a human figure, and the patient is asked to shade in the areas of pain.

The BPI demonstrates a number of advantages for the measurement of cancer pain. The BPI is short and can be completed by most patients in two or three minutes. Depending on the patient, it can be self-administered, given in a clinical interview, or administered over the telephone. The BPI has demonstrated good test-retest reliability over short intervals. Evidence for the validity of the BPI comes from multiple studies of cancer patients (Cleeland, 1989, 1991). For example, patients with metastatic disease reported more severe pain and greater interference in daily activities due to pain than patients without metastatic disease.

Another advantage of the BPI is its availability in multiple languages. The BPI has been administered to cancer patients in numerous countries in Europe, Central and South America, Asia, and the Middle East. The results of these studies suggest that cancer patients in pain from very different cultural and linguistic groups rate pain severity and pain interference in a similar fashion (Cleeland et al., 1996, 1997). Factor analyses of patients' BPI responses reveal two similar factors, pain severity and pain interference, that are remarkably consistent across cultures and languages.

Other pain assessment measures that have been used with cancer patients include the McGill Pain Questionnaire and the Memorial Pain Assessment Card. The original McGill Pain Questionnaire (MPQ) is too long for routine clinical use with cancer patients. The short form of the MPQ (SF-MPQ) consists of 11 sensory (e.g., sharp, shooting) and 4 affective (e.g., sickening, fearful) verbal descriptors (Melzack, 1987). The patient is asked to rate each descriptor on an intensity scale of 0 ("none") to 3 ("severe"). Three pain scores are calculated from the SF-MPQ: the sensory, affective, and total pain rating index (PRI). In addition, patients rate their present pain intensity on a 0–5 scale and a VAS. The SF-MPQ provides important information about the affective component of pain that is not included in many pain questionnaires. However, some of the verbal descriptors may be difficult for patients to comprehend and may be interpreted differently by patients of differing educational or cultural backgrounds.

The Memorial Pain Assessment Card (MPAC) consists of three visual analog scales that assess pain intensity, pain relief, and mood, and one 8-point verbal rating scale that provides another measure of pain intensity (Fishman et al., 1987). This instrument is short and can be administered in the hospital or clinic setting. Some patients may have difficulty comprehending the VAS and verbal rating scale. A Spanish language version of the MPAC recently was validated for use with cancer patients (Thorpe et al., 1999).

Fatigue Assessment

Fatigue is the most frequently reported symptom of cancer patients and often the symptom that causes the greatest amount of interference with daily activities (Glaus et al., 1996; Richardson, 1995). A study of breast and lung cancer patients revealed that 99% of the patients experienced fatigue (Blesch et al., 1991). In a study of chemotherapy and radiotherapy patients, Irvine and colleagues found that 61% experienced fatigue after treatment (Irvine et al., 1994). The high

prevalence of fatigue across diagnostic and treatment categories underscores the need for accurate and regular assessment.

Despite the frequency of fatigue, this symptom is not routinely measured in most oncology settings. Recent literature reviews on cancer-related fatigue point out the need for better assessment tools for measuring fatigue in cancer patients (Irvine et al., 1991; Richardson, 1995). The general approach to assessing fatigue is based on the conceptualization of fatigue as a subjective experience. Although there is not a generally accepted definition of fatigue, it is a multi-dimensional construct with physical (e.g., weakness), cognitive (decreased attention), and affective (e.g., weariness) components.

Recently, several instruments have been employed for the assessment of fatigue in cancer patients. These include: the Pearson-Byars Fatigue Feeling Checklist (Pearson & Byars, 1956), the Fatigue Assessment Instrument (Schwartz, Jandorf, & Krupp, 1993), the Fatigue Symptom Inventory (Hann et al., 1997), the Piper Fatigue Self-Report Scale (Piper et al., 1989), the Multidimensional Fatigue Inventory (Smets, Garssen, Bonke, & De Haes, 1995), and the Multidimensional Fatigue Symptom Inventory (Stein, Martin, Hann, & Jacobsen, 1998). Some of these instruments are too long for very ill cancer patients to complete and are not feasible for routine clinical screening of cancer patients. Others include English-based expressions that may be difficult to understand or translate to other languages.

Two subscales of the Profile of Mood States have been used to assess fatigue in cancer patients (McNair et al., 1992). These subscales are the POMS-Fatigue and the POMS-Vigor. The POMS-Fatigue consists of the following items: “worn-out,” “listless,” “fatigued,” “exhausted,” “sluggish,” “weary,” and “bushed.” The POMS-Vigor includes these items: “lively,” “active,” “energetic,” “cheerful,” “alert,” “full of pep,” “carefree,” and “vigorous.” On both subscales, each item is rated using a 5-point scale. These scales are brief enough to assess fatigue in cancer patients, but may be difficult to translate. Also, patients may interpret the items in various ways, depending on their educational and cultural backgrounds.

The Functional Assessment of Cancer Therapy-Fatigue was designed to measure the fatigue symptoms of cancer patients with anemia (Yellen et al., 1997). It consists of the 28 items of the Functional Assessment of Cancer Therapy-General (FACT-G) to assess health-related quality of life and an additional 13 items to assess fatigue. Examples of these fatigue items are: “I feel weak all over” and “I am frustrated by being too tired to do the things I want to do.” Each item is rated on a 5-point Likert scale ranging from “0” (not at all) to “4” (very much so). The FACT -F demonstrates good internal consistency and test-retest reliability. The main disadvantage of the FACT -F is its length. It may be too long to be given to cancer patients in a clinical setting. The 13-item Fatigue Subscale, on the other hand, does meet the requirement of a rapidly administered scale, but some items may be difficult to understand or to translate into other languages.

The Brief Fatigue Inventory (BFI) was developed to address some of the concerns with existing instruments (Mendoza et al., 1999). Like the MDASI and the BPI, the simple wording of the BFI makes it easy to understand for educationally disadvantaged patients as well as easy to translate. The one-page BFI has only nine items, with the items measured on 0–10 numeric rating scales. Three items ask patients to rate the severity of their fatigue at its “worst,” “usual,” and “now”

during normal waking hours, with 0 being "no fatigue," and 10 being fatigue "as bad as you can imagine." Six items assess the amount that fatigue has interfered with different aspects of the patient's life during the past 24 hours. Depending upon the purpose of measurement, this time interval can be changed to the past week.

The BFI meets standard criteria for reliability and validity, and identifies those with severe fatigue. In the initial validation study of the BFI, 305 adult inpatients and outpatients with cancer completed the BFI and other measures of fatigue (Mendoza et al., 1999). The same instruments were also administered to community-dwelling adults to obtain a comparison sample. The BFI was shown to be an internally stable measure that tapped a single dimension, the subjective report of fatigue severity. The total score on the BFI correlated highly with scores on similar fatigue measures. More than 98% of cancer patients were able to complete the BFI.

Evidence of the validity of the BFI also has been found. A large proportion of the healthy control subjects reported scores at the lower end of the distribution of BFI scores, while the scores of cancer patients tended to be uniformly distributed. A greater proportion of patients reported high levels of fatigue when compared to control subjects. Treatment- or disease-related anemia (represented by hemoglobin level) was significantly associated with BFI scores. In addition, patients with more severe disease, as demonstrated by performance status ratings, also reported greater levels of fatigue than patients with less severe disease.

Fatigue severity ratings on the BFI can identify two groups of patients, those with "severe" and "non-severe" fatigue. Patients who rated their worst fatigue at a 7 or greater demonstrated "severe" fatigue that interfered with multiple areas of daily functioning (Mendoza et al., 1999). Using this range (7–10) to indicate severe fatigue, 35% of the sample of cancer patients, and only 5% of the sample of community controls, were identified as having severe fatigue. Patients who rate their fatigue severity at a 7 or greater can benefit from some type of clinical intervention to improve their function.

INNOVATIVE TRENDS IN SYMPTOM ASSESSMENT

One problem that can occur in using paper-and-pencil ratings of symptoms is that, especially over the life of a relatively long clinical trial, patients forget to fill in the assessment resulting in missing data. Recent developments in computer and communications technology offer new opportunities for the assessment of patients' symptoms. Computer administered versions of symptom ratings scales have been used successfully to screen for psychiatric symptoms. The computer versions of scales appear to be reliable, valid instruments that provide data equivalent to interview or self-administered scales (Kobak et al., 1996). Hand-held computers and other electronic recording devices have been used for the assessment of pain in patients' home and work environments (Lewis, Lewis, & Cumming, 1995). The automated recording devices can be programmed to remind patients to record their symptoms at specified times. Given that memory for pain and other symptoms is often poor, the "real time" assessment of symptoms can provide accurate data regarding symptom patterns and changes over time. However, not all patients are comfortable using hand-held computers or other small automated devices. In addition, patients have to remember to carry and use the devices every day and to bring or transmit the symptom data to their health care providers.

The development of telephone interactive voice response (IVR) technology provides an exciting option for two-way communication with the provider that is acceptable to most patients. Telephone systems have been widely used in outpatient health care settings for communicating with patients, identifying symptoms that need medical attention, and following patients after treatment (Korcz et al., 1998). However, traditional telephone communication requires considerable staff time and is not feasible for assessing symptoms on a regular basis. Using IVR technology that combines touch tone telephones with computers and the internet may be an effective way to follow patients who have symptoms like pain that need to be monitored closely while away from the clinic or hospital. A patient can respond to spoken instructions by using the keypad of a touch tone phone. For example, a patient might be asked to rate his/her pain at its worst in the last day from 0 (no pain) to 10 (pain as bad as you can imagine). Information obtained in this way can be used to update a patient file on an internet or intranet site. In clinical use, the system can be configured to alert physicians and other health care providers. In a pilot study at M.D. Anderson Cancer Center, the system paged a provider when patients reported that the severity of their worst pain in the last 24 hours was 7 or greater (Chandler et al., 1997).

Both the MDASI and the BPI work well in the IVR format. Data obtained from the IVR and paper-and-pencil versions approach equivalency (Mendoza et al., 2001). Even very ill patients are able to complete the IVR assessments and find the assessment burden minimal. Since symptoms vary over time due to treatment or disease effects, repeat IVR assessments help by giving a longitudinal picture of symptom severity. In outcome studies, multiple assessments obtained via IVR allow for analysis of area under the curve (AUC) of symptom severity over time (Wang et al., 2001).

An IVR system should be especially helpful for assessing symptoms such as pain, depression, or sleep disturbance that patients may be reluctant to report to their busy treatment team. Accurate and regular symptom assessment, with data provided to the patient's physicians, should facilitate symptom management. The IVR system also can provide an innovative means of assessing patients symptoms in their home and work environments.

CONCLUSION

Many CAM-based treatments have symptom relief as a target. As with more traditional pharmacologic treatment, it is important to document the effectiveness of these treatments on symptom severity, symptom interference, and patient function. Especially when sick patients are studied, the assessment needs to present minimum patient burden in addition to meeting the expected requirements of validity, reliability and responsiveness. Several single and multiple symptom assessment measures are now available that meet these requirement. Because of our imprecise understanding of the targets of many symptom-oriented interventions, multiple symptom assessment measures should be encouraged. New technologies allow for careful monitoring of patients with reduced missing data and increased patient and researcher convenience.

REFERENCES

Andrykowski MA, Greiner CB, Altmaier EM, Burish TG, Antin JH, Gingrich R, et al. 1995. Quality of life following bone marrow transplantation: findings from a multicentre study. Br J Cancer. 71(6):1322-1329.

Ashbury FD, Findlay H, Reynolds B, McKerracher K. 1998. A Canadian survey of cancer patients' experiences: are their needs being met? J Pain Symptom Manage. 16(5):298-306.

Bergner M, Bobbitt RA, Carter WB, Gilson BS. 1981. The Sickness Impact Profile: Development and final revision of a health status measure. Medical Care. 19:787-806.

Blesch KS, Paice JA, Wickham R, Harte N, Schnoor DK, Purl S, et al. 1991. Correlates of fatigue in people with breast or lung cancer. Oncol Nurs Forum. 18(1):81-87.

Brescia FJ, Adler D, Gray G, Ryan MA, Cimino J, Mamtani R. 1990. Hospitalized advanced cancer patients: a profile. J Pain Symptom Manage. 5(4):221-227.

Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. 1991. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care. 7:6-9.

Cella DF. 1996. Quality of outcomes: Measurement and validation. Oncology. 10:233-244.

Chandler SW, Payne R. 1997. Computerized tools to assess and manage cancer pain. Highlights in Oncology Practice. 14(4):114-117.

Chang VT, Hwang SS, Corpion C, Feuerman M. 1997. Validation of the Memorial Symptom Assessment Scale Short Form (MSAS-SF). Proc Am Soc Clin Oncol. 16:47a [Abstract].

Cleeland CS. 1989. Measurement of pain by subjective report. In C.R. Chapman and J.D. Loeser (Eds.), Issues in Pain Measurement. New York: Raven Press.

Cleeland CS. 1990. Assessment of pain in cancer: Measurement issues. In K.M. Foley, J.J. Bonica, V. Ventafridda, and M.V. Callaway (Eds), Advances in Pain Research and Therapy (Vol. 16, pp. 47-55). New York: Raven Press.

Cleeland CS. 1991. Research in cancer pain: What we know and what we need to know. Cancer. 67(Suppl. 3):823-827.

Cleeland CS, Gonin R, Hatfield AK, Edmonson JH, Blum RH, Stewart JA, Pandya KJ. 1994. Pain and its treatment in outpatients with metastatic cancer. New England Journal of Medicine. 330:592-594.

Cleeland CS, Nakamura Y, Mendoza TR, Edwards KR, Douglas J, Serlin RC. 1996. Dimensions of the impact of cancer pain in a four country sample: New information from multidimensional scaling. Pain. 67(2-3):267-273.

Cleeland CS, Gonin R, Baez L, Loehrer P, Pandya KJ. 1997. Pain and treatment of pain in minority outpatients with cancer. The Eastern Cooperative Oncology Group minority outpatient pain study. Annals of Internal Medicine. 127:813-816.

Cleeland CS. 1999. Measuring and Understanding Fatigue. NCCN Proceedings, Oncology. 13:(11A).

Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, Engstrom MC. 2000. Assessing symptom distress in cancer patients: the M. D. Anderson Symptom Inventory. Cancer. 89(7):1634-1646.

Coyle N, Adelhardt J, Foley KM, Portenoy RK. 1990. Character of terminal illness in the advanced cancer patient: pain and other symptoms during the last four weeks of life. J Pain Symptom Manage. 5(2):83-93.

de Haes JC, van Knippenberg FC, Neijt JP. 1990. Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. Br J Cancer. 62(6):1034-1038.

Degner LF, Sloan JA. 1995. Symptom distress in newly diagnosed ambulatory cancer patients and as a predictor of survival in lung cancer. J Pain Symptom Manage. 10(6):423-431.

Fainsinger R, Miller MJ, Bruera E, Hanson J, Maceachern T. 1991. Symptom control during the last week of life on a palliative care unit. J Palliat Care. 7(1):5-11.

Fairfield KM, Eisenberg DM, Davis RB, Libman H, Phillips RS. 1998. Patterns of use, expenditures, and perceived efficacy of complementary and alternative therapies in HIV-infected patients. Arch Intern Med. 158(20):2257-2264.

Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. 1990. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. Journal of Rheumatology. 17:1022-1024.

Fishman B, Pasternak S, Wallenstein SL, Houde RW, Holland JC, Foley KM. 1987. The Memorial Pain Assessment Card. A valid instrument for the evaluation of cancer pain. Cancer. 60:1151-1158.

Glaus A, Crow R, Hammond S. 1996. A qualitative study to explore the concept of fatigue/tiredness in cancer patients and in healthy individuals. Support Care Cancer. 4(2):82-96.

Gonzales GR, Elliott KJ, Portenoy RK, Foley KM. 1991. The impact of a comprehensive evaluation in the management of cancer pain. Pain. 47:141-144.

Harrison LB, Zelefsky MJ, Pfister DG, et al. 1997. Detailed quality of life assessment in patients treated with primary radiotherapy for squamous cell cancer of the base of the tongue [see comments]. Head Neck. 19:169-175.

Hickok JT, Morrow GR, McDonald S, Bell AJ. 1996. Frequency and correlates of fatigue in lung cancer patients receiving radiation therapy: implications for management. J Pain Symptom Manage. 11(6):370-377.

Ingham JM, Portenoy RK. 1996. Symptom assessment. Hematol Oncol Clin North Am. 10:21-39.

Irvine D, Vincent L, Graydon JE, Bubela N, Thompson L. 1994. The prevalence and correlates of fatigue in patients receiving treatment with chemotherapy and radiotherapy. A comparison with the fatigue experienced by healthy individuals. Cancer Nursing. 17(5):367-378.

Irvine DM, Vincent L, Bubela N, Thompson L, Graydon J. 1991. A critical appraisal of the research literature investigating fatigue in the individual with cancer. Cancer Nurs.14(4):188-199.

Jacobsen PB, Hann DM, Azzarello LM, Horton J, Balducci L, Lyman GH. 1999. Fatigue in women receiving adjuvant chemotherapy for breast cancer: characteristics, course, and correlates. J Pain Symptom Manage. 18(4):233-242.

Jacox A, Carr DB, Payne R, Berde CB, Breitbart W, Cain JM, Chapman CR, et al. 1994. Management of cancer pain. Clinical practice guideline No. 9. AHCPR Publication no. 94-0592. Rockville, MD: Agency for Health Care Policy and Research, U.S. Department of Health and Human Services, Public Health Service.

Jenkins CA, Taube AW, Turner K, Hanson J, Bruera E. 1998. Initial demographic, symptom, and medication profiles in patients admitted to continuing palliative care units. J Pain Symptom Manage. 16:163-70.

Jensen MP, Karoly P, Braver S. 1986. The measurement of clinical pain intensity: a comparison of six methods. Pain. 27:117-126.

Kobak KA, Greist JH, Jefferson JW, Katzelnick DJ. 1996. Computer-administered clinical rating scales. A review. Psychopharmacology (Berl). 127(4):291-301.

Kornblith AS, Thaler HT, Wong G, et al. 1995. Quality of life of women with ovarian cancer. Gynecol Oncol. 59:231-42.

Korcz IR, Moreland S. 1998. Telephone Prescreening Enhancing a Model for Proactive Healthcare Practice. Cancer Pract. 6(5):270-275.

Lewis B, Lewis D, Cumming G. 1995. Frequent measurement of chronic pain: an electronic diary and empirical findings. Pain. 60:341-347.

McCorkle R, Young K. 1978. Development of a symptom distress scale. Cancer Nurs. 1(5):373-378.

McCorkle R, Quint-Benoliel J. 1983. Symptom distress, current concerns and mood disturbance after diagnosis of life-threatening disease. Soc Sci Med. 17(7):431-438.

McCorkle R. 1987. The measurement of symptom distress. Semin Oncol Nurs. 3(4):248-256.

McNair DM, Lorr M, Droppleman LF. 1992. EdITS Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service.

McQuellon RP, Craven B, Russell GB, Hoffman S, Cruz JM, Perry JJ, et al. 1996. Quality of life in breast cancer patients before and after autologous bone marrow transplantation. Bone Marrow Transplant. 18(3):579-584.

Melzack R. 1987. The short-form McGill Pain Questionnaire. Pain. 30:191–197.

Mendoza TR, Chou C, Cleeland CS. 1999. The relationship of pain to the most commonly reported symptoms associated with cancer. Presented at the 9th World Congress on Pain, Vienna, Austria.

Mendoza T, Anderson K, Wang XS, Easley M, Mobley G, Cleeland CS. 2003. Phone it in or fill it in? Comparing symptom assessment using the interactive voice response system with the traditional paper and pencil method. Submitted to QOL Research.

Moinpour CM. 1994. Measuring quality of life: An emerging science. Seminars in Oncology. 21:48-60.

Moinpour CM, Feigl P, Metch B, Hayden KA, Meyskens FL Jr, Crowley J. 1989. Quality of life end points in cancer clinical trials: Review and recommendations. Journal of the National Cancer Institute. 81:485-495.

Philip J, Smith WB, Craft P, Lickiss N. 1998. Concurrent validity of the modified Edmonton Symptom Assessment System with the Rotterdam Symptom Checklist and the Brief Pain Inventory [In Process Citation]. Support Care Cancer. 6:539-41

Portenoy RK, Thaler HT, Kornblith AS, et al. 1994. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer. 30A:1326-36.

Portenoy RK, Kornblith AS, Wong G, et al. 1994. Pain in ovarian cancer patients. Prevalence, characteristics, and associated symptoms. Cancer. 74:907-15.

Prieto JM, Saez R, Carreras E, Atala J, Sierra J, Rovira M, et al. 1996. Physical and psychosocial functioning of 117 survivors of bone marrow transplantation. Bone Marrow Transplant. 17(6):1133-1142.

Richardson A. 1995. Fatigue in cancer patients: a review of the literature. Eur J Cancer Care (Engl). 4(1):20-32.

Samarel N, Leddy SK, Greco K, Cooley ME, Torres SC, Tulman L, Fawcett J. 1996. Development and Testing of the symptom experience scale. J Pain Symptom Manage. 12(4):221-228.

Schuit KW, Sleijfer DT, Meijler WJ, et al. 1998. Symptoms and functional status of patients with disseminated cancer visiting outpatient departments. J Pain Symptom Manage.16:290-7.

Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. 1995. When is cancer pain mild moderate or severe? Grading pain severity by its interference with function. Pain. 61:277-284.

Smets EM, Garssen B, Cull A, de Haes JC. 1996. Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. British Journal of Cancer. 73(2):241-245.

Spilker B. 1996. Quality of life and pharmacoeconomics in clinical trials. Philadelphia: Lippincott-Raven.

Syrjala KL. 1987. The Measurement of Pain. In: Yarbro CH, McGuire DB, editors. Cancer Pain: Nursing Management. Orlando: Grune and Stratton, Inc.; pp. 133-150.

Trowbridge R, Dugan W, Jay SJ, Littrell D, Casebeer LL, Edgerton S, Anderson J, O'Toole JB. 1997. Determining the effectiveness of a clinical-practice intervention in improving the control of pain in outpatients with cancer. Academic Medicine. 72:798-800.

Vainio A, Auvinen A. 1996. Prevalence of symptoms with advanced cancer: An international collaborative study. Journal of Pain and Symptom Management. 12:3-10.

Von Roenn JH, Cleeland CS, Gonin R, Hatfield AK, Pandya KJ. 1993. Physician attitudes and practice in cancer pain management. A survey from the Eastern Cooperative Oncology Group. Annals of Internal Medicine. 119:121-126.

Wang XS, Cleeland CS, Mendoza TR, Engstrom MC, Liu S, Xu G, Hao X, Wang Y, Ren XS. 1999. The effects of pain severity on health-related quality of life. Cancer. 86:1848-1855.

Wang XS, Janjan N, Mendoza T, Johnson B, Crane CH, Engstrom MC, Cleeland CS. 2001. Fatigue During Preoperative Chemoradiation for Resectable Rectal Cancer. Cancer (Supplement). 92(6):1657-1744.

Ward SE, Goldberg N, Miller-McCauley V, Mueller C, Nolan A, Pawlik-Plank D, Robbins A, Stormoen D, Weissman DE. 1993. Patient-related barriers to management of cancer pain. Pain. 52:319-324.

Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. 1997. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage. 13(2):63-74.

THEORY, HYPOTHESIS DEVELOPMENT, AND RANDOMIZED CLINICAL TRIALS IN COMPLEMENTARY AND ALTERNATIVE MEDICINE; A BEGINNING AT SEPARATING SCIENCE FROM POLITICAL SCIENCE

Gary R. Morrow, PhD*
Professor of Radiation Oncology and Psychiatry
University of Rochester Cancer Center
Behavioral Medicine Unit

Jane T. Hickok, MD, MPH
Research Assistant Professor
Oncology in Preventive and Community Medicine
University of Rochester School of Medicine and Dentistry

Joseph A. Roscoe, PhD
Research Assistant Professor
Radiation Psychology
University of Rochester Cancer Center

Sara E. Matteson, PsyD
Postdoctoral Research Associate
University of Rochester Cancer Center

Cancer Control Program, University of Rochester Cancer Center, Rochester, New York

**Corresponding author:*
Gary R. Morrow, PhD
University of Rochester Cancer Center
Behavioral Medicine Unit
601 Elmwood Avenue, Box 704
Rochester, New York 14642
Tel: (716) 275-5513
Fax: (716) 461-5601
E-mail: Gary _Morrow@URMC.Rochester.edu

THEORY, HYPOTHESIS DEVELOPMENT, AND RANDOMIZED CLINICAL TRIALS IN COMPLEMENTARY AND ALTERNATIVE MEDICINE; A BEGINNING AT SEPARATING SCIENCE FROM POLITICAL SCIENCE

INTRODUCTION:

There is growing recognition that many people use health and wellness interventions commonly included under the term Complementary and Alternative Medicine (CAM).[1] Along with this awareness have come questions about the worth and effectiveness of CAM procedures—questions that call for evidence-based answers. As CAM research increases, issues have been raised related to the adequacy of current theory, the development of new theoretical frameworks for investigating CAM modalities, and the implications of subjecting CAM to the rigors of evidence-based investigation. The purpose of this paper is to begin a discussion of some of these important issues with a view to focusing debate and encouraging further dialog. A summary of what constitutes a testable theory and a reasonable hypothesis is outlined as a prelude to a discussion of design aspects of CAM research. Following a discussion of the importance of non-specific effects and their relative importance in CAM as well as traditional medical research, a brief history of the development of the various phases of randomized clinical trials is presented along with the suggestion that these phases can be viewed as resulting as much from political and policy forces as from methodological and scientific ones.

A BRIEF PRIMER ON BASICS:

A theory is an organized, integrated and explanatory set of principles concerning a phenomenon or observation. A theory allows for predictions of events and circumstances, unifies a set of ideas, and illustrates how these components are thought to relate to one another.[2,3] Theories of health and health care are continually being proposed to foster understanding of disease and determine the effects of potential interventions. For example, there are genetic and psychological theories regarding the development of anorexia and bulimia nervosa, as well as biochemical/ neuro-immunological theories to explain conditions such as Parkinson's disease and schizophrenia. In addition, there are numerous theories about potential causes of Alzheimer's disease. As an illustration of the latter, some say an imbalance of neurotransmitters, especially acetylcholine, is the culprit. Others extend this thinking, citing a slow-acting virus that destroys brain cells as the source of the acetylcholine deficiency, while still others attribute the disease to brain trauma.[4]

Several criteria determine how “good” a scientific theory is. First, the theory should correspond with accepted knowledge rather than clash with it. Of course, what is initially accepted as “accepted knowledge” might eventually be proven wrong, which relates to the second characteristic of a good theory—it must be falsifiable (i.e., refutable). As stated over 70 years ago by Karl Popper, a philosopher and scientific theorist, anyone with a fertile imagination can find evidence to support almost any imaginable theory. And the history of medicine is replete with fanciful theories that were accepted as dogma in their time. The key to a valid scientific theory, Popper proposed, is that there must be a way to empirically disprove the correctness of a theory, rather than accepting it on faith. Accordingly, the theory that humans are direct

descendants of extraterrestrial beings is not a good scientific theory because it cannot be refuted by any conceivable observation.[5] Third, a good theory is coherent. The ideas contained in it "hang" together well, and do so in a simple, straightforward way. In addition, useful theories are parsimonious and direct rather than complicated and superfluous.[6]

In any field, research involves testing theories to determine their accuracy and gather information to create new avenues of exploration. A first step in any such endeavor is the development of a hypothesis, or testable prediction derived from the theory. While theories are broad and have a large focus, hypotheses are narrow in scope.[6] Hypotheses attempt to explain and clarify the relationship between two or more variables or phenomena, with an "if-then" format: the "if" portion refers to the independent variable, or that which can be altered or manipulated at will. The "then" component refers to the dependent variable, or the resulting data. For example, cognitive theory stipulates that a pattern of negative thinking in which one employs cognitive distortions can lead to and maintain depression. The testable hypothesis, then, would be that if an individual learns to challenge these distortions (i.e., to stop disqualifying positive events, to begin placing responsibility for negative outcomes where it belongs rather than on oneself), then the depression would improve. To test the hypothesis, the phenomena in question are operationally defined, the exact nature of what is to be observed or measured is outlined. To continue the above example, the number of negative cognitive distortions per day might first be recorded for a given length of time, and a measure of depression, such as the Beck Depression Inventory, obtained. After treatment designed to teach the individual to challenge the negative thoughts with positive statements, the frequency of distortions and level of depression would be assessed again. If a measurable improvement in the outcomes has occurred, the resulting data would be taken as confirmation of the hypothesis if no alternative explanation can account for the results.[2]

Over the years, a number of disease theories have been proposed and tested in such a manner, with a mixture of results. For example, there is an apparent excess of winter-time birth dates among people who develop schizophrenia, and a deficiency of summer-time births.[7,8] Numerous explanations for this phenomenon have been proposed, including variations in the procreation patterns of the parents of affected individuals, and the effect of seasonal temperature and diet changes. There was also an age-prevalence explanation suggesting that since those born in the winter were also born earlier in the year, and thus older, they were at greater risk. Since age increases the risk of the condition (to some degree), it was stated that those born later in the year were simply younger and therefore didn't have the same risk potential.[8] Exposure to infectious agents was also considered, and research has found evidence for this theory. Seasonal variation occurs with regard to the viral agents that are active, and individuals born in the winter are more likely to be exposed to viral infections than those born in warmer months.[7]

The history of schizophrenia also provides evidence of the dangers of not basing treatments on scientific knowledge. In the 19th century, afflicted individuals were subjected to insulin-induced comas, bilateral electro-convulsive therapy (ECT) and pre-frontal lobotomy. These methods were employed without a full understanding of the mechanisms that led to the observed behaviors and, more importantly, without a true appreciation for their impact on future health and behavior. Thus, well-thought out theories, and tests of them, are extremely important.

Another illustrative example of the importance of hypotheses and theory occurred in the early months of the AIDS epidemic. When doctors first noticed unusual symptoms and conditions in their young, male patients, numerous notions about the cause were suggested. Some related the condition to parasitic diseases associated with the sexual practices of gay men. Others said it was caused by some toxic substance that was in the environment of these individuals, and that the occurrences of the strange illnesses would fade away as they had begun.[9] Had these ideas simply been taken at face value, and rigorous methods not employed to discover the blood-borne virus that compromised the immune system, the prevalence of infection would be even greater today. As these examples demonstrate, it is important that sound methods be implemented when conducting research, especially when the results can greatly impact human lives. Additional crucial features of any research study that must be considered when doing CAM research concern the reliability and validity of the outcome measures being used. Reliability reflects the consistency of a measure. In other words, a test is reliable when it yields reproducible, dependable, and stable results.[2,10]

In classical psychometric theory, reliability also has a more specific meaning reflecting the amount of precision and the amount of error in a test score. Any score on a test has two parts: one reflects the true score, or stable factor that is measured, such as intelligence, anxiety, etc. The other part reflects error, the random noise that is reflected in the score which can be affected by guessing, fatigue, motivation, and the like.[10] A perfectly reliable test has no error at all, whereas a perfectly unreliable test tells nothing about the trait or attribute being measured.

Validity refers to the extent to which a test measures what it is supposed to measure and the degree to which inferences can be drawn on the basis of results.[10] Types of validity include construct validity, which is the degree to which the measure reflects the concept of interest, and criterion validity which relates to how well a measure correlates with some other variable or outcome. Criterion validity includes convergence, i.e., high correlation with related concepts, and discrimination, i.e., low correlation with unrelated concepts.[2]

The concepts of reliability and validity are related, in that for a test to be valid, it must be reliable. However, high reliability does not ensure validity.[3] A measure might consistently yield the same results, but that doesn't guarantee accuracy of the information obtained. Outcome measures that are considered valid and reliable when used in more conventional avenues of clinical research may or may not be appropriate for use in studies measuring the impact of CAM interventions. This issue must be considered during the design of CAM studies.

DESIGN ISSUES IN CAM STUDIES:

A theoretical understanding of what the concept of healing encompasses best precedes a discussion of what constitutes a well-designed approach to the scientific study of CAM modalities. Kleijnen and colleagues[11] propose that the healing process comprises three disparate elements: 1) the self-healing properties of the body, 2) specific effects of physical or pharmacological interventions, and 3) changes induced by nonspecific effects of the treatment. The term, “nonspecific effect,” is often used synonymously with “placebo effect” and includes aspects of treatment, such as the setting in which the therapy takes place, physician attention and concern, patient and physician expectations of treatment effects, and the credibility of both the

therapist and the treatment.[12] Many CAM practitioners would likely include in this list the intent and/or consciousness of both healer and healee.[13]

Understanding and taking into account the distinction between specific and non-specific factors in healing is critically important in the design of research examining treatment efficacy. Specific effects of a treatment can usually be isolated and measured with both accuracy and efficiency in classically designed experiments, in which: 1) double-blind procedures are used to eliminate potential bias by experimenters and participants, 2) participants are randomized to control for individual variation in the natural course of the disease under investigation and self-healing, and 3) subjects not receiving the actual treatment receive a placebo treatment in order to hold nonspecific factors between treatment and non-treatment groups constant.

What is commonly called Western medicine is focused on specific effects of treatment and for this reason considers the double blind, randomized placebo-controlled trial to be the gold standard in research design.[14] For a treatment to be deemed effective, it must demonstrate efficacy above and beyond that which can be accounted for by non-specific treatment effects, i.e., the treatment group must have a better response than the placebo control group.

CAM practitioners, on the other hand, generally take a more holistic approach to the healing process and intentionally include non-specific treatment effects such as caring and attention as integral components of their therapies.

Herein lies the dilemma: Western medicine researchers treat non-specific effects as something to be controlled for while CAM practitioners may consider them essential elements of treatment. What one field considers extraneous "noise" to be eliminated from any controlled investigation in order to get cleanly to the "active ingredient," may very well be the "active ingredient" of the other field. The fact that many CAM therapies rely heavily on non-specific treatment effects is not a problem in the clinical practice of CAM, because any treatment effect, whether it be specific or non-specific, that is beneficial to patients is clearly a good thing. (And it is likely also important for Western medicine).

These effects, however, may well be a significant obstacle to conducting CAM research because of the pronounced difficulty in conducting scientific research when non-specific effects are included as part of the research design rather than as factors to be controlled. Non-specific effects can cloud interpretation: it is nearly impossible to separate specific from non-specific treatment effects, thereby leading to problems with answering such basic research questions as "was a specific treatment effect present, and if so, what is its magnitude?" A second difficulty is in the control of experimenter artifacts, particularly the elimination of bias. Research outcomes that can be influenced by intangible factors such as patients' and providers' treatment expectancies and attitude can by their very nature also be affected by patients' and providers' biases.

Both problems are related to the fact that it is difficult to isolate and/or measure the effects caused by individual non-specific factors on treatment outcome. Non-specific factors are frequently numerous and likely to interact with one another. Researchers who choose to isolate and manipulate a single non-specific factor may have problems with experimental power because

of small effect sizes. On the other hand, researchers who take the other route and create an experimental condition in which multiple, and potentially interacting, non-specific factors are affected by the intervention, may have problems with interpretation of the results. For example, in a seemingly simple experiment comparing pain relief between subjects receiving a CAM herbal remedy with subjects not receiving the herbal remedy, a variety of nonspecific factors might play an important role in determining the outcome. Some of these factors include: the credibility of the researcher, the nature of the research environment, the size, shape, and taste of the remedy, and differences in subject response due to being randomized into noticeably different experimental treatment conditions (i.e., herbal remedy vs. no herbal remedy). Without a credible control condition (e.g., an identical looking, tasting, and smelling herbal remedy without demonstrated effect) it is not possible to separate the effects of the non-specific factors from the specific factors of treatment.

If lack of clear interpretation of otherwise well-conducted CAM research that showed positive findings was the only problem, we might all happily live with it. After all, the exact mechanism of action of many conventional therapies is not well understood and we accept their validity provided that patients reliably benefit from the treatment.[15] Unfortunately, the above herbal remedy example also has the additional problem of potential participant bias inherent in experiments in which subjects can guess the nature of the experiment and which treatment arm they have been assigned to, i.e., not blinded. Participants in such an experiment, because of their desire to see either the experiment or the experimenter be successful, might consciously or subconsciously be obliging and report results that they think are desired. The lack of adequate patient blinding can be a prominent challenge to CAM research due to the difficulty of providing credible, yet non-therapeutic control conditions. It is important to note that it is the inclusion of placebo-control groups in modem experiments (i.e., those requiring informed consent) that allows for experiments to be blinded, and that is one of the primary reasons that they are used.

A BRIEF HISTORY OF HOW CLINICAL TRIALS DEVELOPED

In his wonderful book "Diffusion of Innovations," Everett Rogers[16] relates the story of British sea captain James Lancaster who tested the effectiveness of lemon juice for scurvy in 1601 while he commanded four ships sailing from England to India. His experimental design was to serve three teaspoons of lemon juice to the crew on one of the ships. Most of the crew on the experimental ship remained healthy. On the other three ships 110 of 278 died of scurvy—an undesirable outcome that led to the crew of the experimental ship being scattered to the others so they had crew enough to get safely to port. This event has been little noticed for Lancaster had no credentials: he was neither scientist nor naval surgeon. Findings from this first clinical trial appear to have been ignored for over a century and a half until they were replicated—by an investigator with proper credentials.

In 1747 James Lind, a British naval physician, conducted what is often cited as the first controlled clinical trial. Seamen aboard the HMS Salisbury who had developed scurvy were put in one of five groups, to receive one of the following regimens each day: 1) two oranges and one lemon, 2) a half-pint of seawater, 3) six spoonsful of vinegar, 4) a quart of cider, or 5) 75 drops of a potion called "vitriol elixir." The sailors treated with citrus were cured within a few days; subjects in the other treatment groups were not. Even given the clear results of the trial, and the

scientific and medical credentials of the principal investigator, it still took the British navy another half century to adopt the findings. Not until 1795 did the British navy finally adopt the regular shipboard use of citrus fruit; potentially lethal scurvy was eradicated almost overnight. This vignette highlights some important aspects of the intermix of science and political science that may have application to CAM research using clinical trials methodology. The first parallel is that investigators with "proper" credentials are often given more gravitas than investigators who do not have mainstream research backgrounds.

A second common lesson is the time lag between research findings and their adoption into clinical practice. The current emphasis on "translational research" is a direct-line descendent of an earlier call for "diffusion" of results from bench to bedside (which seemed to drop out of favor since it sounded a bit too much like chemical engineering). "Diffusion" evolved over time into the need for "technology transfer" of findings from bench to bedside (which also faded from popular lexicon with revelations of five-hundred-dollar NASA hammers and thousand-dollar technologically sophisticated toilets). The point is that clinical research has a timeline of its own that seems impervious to popular or political pressure.

And so it was with the citrus and scurvy research story. It unfortunately didn't end with the unconscionable delay of 194 years from first data until the intervention was adopted by the British navy (or the equally appalling 49 years from confirmatory clinical trial data from one of their own respected investigators). The British Board of Trade did not adopt citrus fruits as prevention for scurvy in the merchant marine until 1865 with the same eradication of scurvy as noted 70 years earlier in the Royal Navy. From the availability of initial clinical trial data to the change in clinical practice was a startling two and a half centuries. Hopefully, such patience will not be required for CAM research findings to reach clinical practice.

THE ROLE OF THE U.S. GOVERNMENT IN CLINICAL RESEARCH:

Until shortly before the second world war, the U.S. government did not regulate pharmaceutical agents or exercise oversight in their development or distribution. In 1938, after years of debate among lawmakers about the necessity for such a law,[17,18] Congress passed the Food, Drug, and Cosmetic Act. The Act was spurred into passage by the occurrence of more than 100 deaths in persons taking a particular formulation of a new drug, "Elixir Sulfanilamide," for infections. (In fact, it was later shown that the solvent, diethylene glycol, not the drug, was the culprit.)

While this legislation did not require assessment of efficacy, it gave the FDA authority to review the safety and composition of new drugs before a drug reached the market. Potential toxicity for human use was the primary concern; documentation of animal testing for toxicological effects was required.[19,20] However, there were no formal criteria for determination of safety, and all assessments were made using data and information supplied by researchers and reports from manufacturers without the FDA defining the type of investigations that had to be performed.

Despite this new authority, the FDA at that time was not authorized to "approve" new drugs; its power resided in its ability to refuse to permit a new drug application to become effective.[21] Although there was no requirement in the law that efficacy be proven, and the FDA lacked the authority to require proof of efficacy, there is some evidence that FDA commissioners at the time

thought that an assessment of safety required a concurrent assessment of therapeutic effectiveness based on clinical evidence from respected researchers.[21]

In 1960, the National Academy of Sciences recommended that the FDA should be given the authority to require proof of efficacy as well as safety of new drugs. However, it was not until 1962, following the documentation of birth defects in children born to mothers in Western Europe who took the sedative Thalidomide while pregnant, that the U.S. Congress passed the Kefauver-Harris Drug Amendment. This legislation mandated that manufacturers prove the effectiveness of drugs as well as their safety before they were released to the U.S. market by the FDA. It also mandated specific FDA approval of new drugs. Efficacy and safety were to be shown through "adequate and well-controlled" studies which became the standard for documenting drug effectiveness. Reporting of all adverse effects was also mandated by the 1962 law, and for the first time patients were required to give informed consent before participating in clinical trials of new drugs.[19] Under FDA auspices, new drug testing progresses through a series of well-characterized clinical trials designed to determine the safety and efficacy of a potential new drug. Phase I, II and III trials have specific purposes and are characterized by specific design requirements. Figure 1 below outlines this progression.

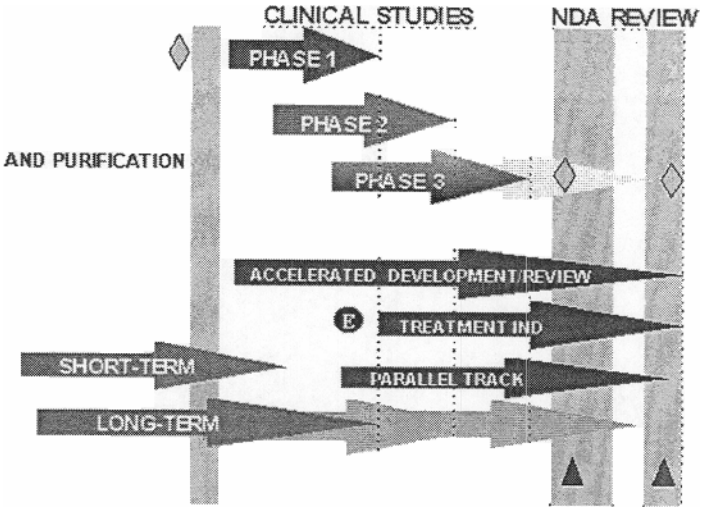


Figure 1. The New Drug Development Process: Steps from Test Tube to New Drug Application Review[21]

Although there can be overlap of intent, the progression among the phases are commonly regarded as: 1) Phase I studies enroll a small number of healthy or ill subjects to gather preliminary data on a new drug's safety. 2) Phase II trials determine most-effective-dose level for efficacy and continue to evaluate the drug's safety. 3) Phase III studies evaluate the effectiveness of the new drug in large numbers of patients with a particular disease or condition who are randomly assigned to the new drug or to the current standard treatment, or to a placebo if no standard treatment exists. Safety also continues to be assessed in Phase III trials.

The history of how this procedure of Phases developed is obscure. A Medline search of the medical literature and a search of official internet sites maintained by the FDA and the NCI for information on the history of clinical trials failed to yield a document that specifically described

the manner by which the series of Phase studies currently required by the FDA for the process of new drug approval came to be established. Dr. Michael Weintraub, former director of the FDA's Office of Drug Evaluation in the Division of Over the Counter Drug Products, suggested that many agency rules had come about as the result of accidents, emergencies, and tragedies such as those that led to the passage of the 1938 law and 1962 Amendment. He speculated that similar circumstances might have led to the adoption of the strict criteria for determining drug safety and efficacy (personal communication, September, 2001). As part of the implementation of the 1962 Amendments, Arthur D. Little, Inc. submitted suggestions for the classification of clinical trials, as well as for a comprehensive system for collecting reports of adverse drug reactions, and in 1966, the Pharmaceutical Manufacturers Association accepted the definitions of Phase I, II and III trials used to monitor the progress of new drug development.[18]

In the years since 1962, the FDA standard for proving efficacy, i.e. "adequate and well-controlled" investigations has been widely adopted as has the requirement for informed consent. The accepted definitions for clinical trial phases have become integral not only to the process of new drug approval, but also to the conduct of clinical research in many, if not most, fields of medical endeavor, notable exceptions being preventive medicine and surgery.[22] Similarly, the requirement for informed consent now includes not only studies of investigational drugs but also most other research studies and medical procedures.[22] Furthermore, the FDA has increased its surveillance of the safety and efficacy of drugs by strongly encouraging, although not outright requiring, the performance of post-marketing studies.[22]

Research in oncology has closely followed the standards promulgated by the FDA and in recent years there has been a shortening of the review time for oncologic drugs. In recent years, development of new classes of therapeutic agents has accelerated, but FDA standards for proving safety and efficacy have remained as they have been for cytotoxic agents. The Washington Clinical Trials Conference held in July 1995 documented significant dissatisfaction with the then-current situation for clinical cancer trials and endorsed a number of recommendations for future trials of non-cytotoxic agents that would provide more timely accumulation of significant clinical and biologic data as well as faster approval of new cancer drugs and other classes of anti-cancer agents.[23]

CONCLUSION

Pre-approval studies of CAM modalities might also need to be examined and altered to reflect their generally less toxic attributes. For example, many interventions that are often included under the CAM rubric, such as T'ai Chi or yoga, or even prayer, are inherently non-toxic and, should not require Phase I toxicity and safety studies prior to being examined in randomized intervention studies. Furthermore, while effects could conceivably be "dose"-related, separate dose-response studies may not be indicated in the absence of toxicity. For CAM modalities that bear closer similarities to conventional Western medicine, for example Chinese herbal mixtures or substances used in traditional Indian medicine, Phase I toxicity studies may be indicated and required by the FDA prior to approval.

The ideal CAM research design, as in conventional medicine, remains the double-blind, randomized, placebo-controlled trial because both non-specific effects and bias are eliminated. In

the ideal experiment, the placebo-control should be indistinguishable from the active treatment. This model does not fit all CAM modalities. So increased emphasis should be placed on the replication of research results by several independent researchers. While the necessity for study replication is well established in conventional medical research, even of studies using ideal designs, this need assumes even greater importance when experiments are not double-blinded, as in the case of some current CAM research. This is akin to the concept of falsifiability described earlier. CAM modalities, to be scientifically credible, must be open to the possibility of being shown ineffective. To enable this, CAM experiments must be designed so they are replicable, and when published, must be described in sufficient detail to make replication possible.

This basic injunction may be a challenge as many CAM remedies are not standardized, e.g., what exactly is in a Chinese herbal mixture, and aspects of many CAM therapies are not well-described, e.g., what is the sensation of rising Chi in acupuncture? Greater emphasis, precision, and clarity are needed in operationally characterizing any CAM intervention under investigation. "Treatment manuals" are often used in multicenter clinical trials of behavioral interventions. Whenever possible, they should be developed and available for others to use in CAM research.

For experiments that cannot be double-blinded, a three-arm trial design comprising an active treatment arm, a non-treatment control arm, and a credible control arm should be considered. The credible control group needs to receive a treatment in which the non-specific treatment factors are as equivalent to those of the active treatment group as possible. Patients randomized to the credible control treatment group must receive the same amount of care and attention as patients in the active treatment group and must believe that they are receiving an active treatment. Measures of relevant non-specific treatment effects, e.g., expected efficacy of the treatment, perceived caring and concern of treatment providers, etc., should be given to participants and used as covariates in all relevant analyses.

The non-treatment control group should be included in order to detect the presence of non-specific treatment effects, and, if present, to estimate their magnitude. If only non-specific treatment effects are found to be present, further studies may still be warranted to separate experimental artifact from true benefit to patients. A treatment that provides true patient benefit by an unknown mechanism (i.e., non-specific factors) could be viewed as a starting point for additional research.

While most of what is now being called "Complementary and Alternative Medicine" predates what is being called "Western Medicine" by centuries, the need to evaluate both systems using the same standards is rather new. As standards and common methodologies evolve, it is useful to remember that virtually all currently accepted medical interventions were seen as unproven, complementary, and alternative at one time.

In time, science may help move several promising current CAM approaches into mainstream medical acceptance. Showing they are effective by using the common methodological standards outlined here can aid that acceptance.

REFERENCES

1. Wootton JC, Sparber A. Surveys of complementary and alternative medicine: part I. General trends and demographic groups. *Journal of Alternative & Complementary Medicine* 2001;7:195-208.

2. Kazdin AE. Research Design in Clinical Psychology. Boston: Allyn and Bacon, 1992.

3. Meyers DG. Psychology. New York: Worth Publishers, Inc., 1986.

4. Nevid IS, Rathus SA, Greene B. Abnormal Psychology in a Changing World. New York: Prentice Hall, 1994.

5. Popper KR. Logik der Forschung. Vienna: Springer-Verlag, 1934.

6. Rosenthal R, Rosnow RL. Essentials of Behavioral Research: Methods and Data Analysis. New York: McGraw-Hill, Inc., 1991.

7. Birchwood MJ, Hallett SE, Preston MC. Schizophrenia: An Integrated Approach to Research and Treatment. New York: New York University Press, 1989.

8. Hirsch SR, Weinberger DR. Schizophrenia. Cambridge: Blackwell Science, 1995.

9. Shiltz R. And the Band Played On: Politics, People and the AIDS Epidemic. New York: Penguin Books, 1987.

10. Sattler JM. Assessment of Children: Revised and Updated. San Diego: Jerome Sattler Publishers, Inc., 1988.

11. Kleijnen J, de Craen AJ, van Everdingen J, Krol L. Placebo effect in double-blind clinical trials: A review of interactions with medications. *Lancet* 1994;344:1347-1349.

12. Turner JA, Deyo RA, Loeser JD, Yon Korff M, Fordyce WE. The importance of placebo effects in pain treatment and research. *JAMA* 1994;271:1609-1614.

13. Dossey L. How should alternative therapies be evaluated? An examination of fundamentals. *Alternative Therapies in Health & Medicine* 1995;1:6-10.

14. Feinstein AR. Current problems and future challenges in randomized clinical trials. *Circulation* 1984;70:767-774.

15. Dossey L. On double-blinds and double standards: A response to the recent New England Journal Editorial. *Alternative Therapies in Health & Medicine* 1998;4:18-20.

16. Rogers EM. Diffusion of Innovations (Third Edition). New York: Free Press, 1983.

17. Hutt PB. Investigations and reports respecting FDA regulation of new drugs (part II). *Clin Pharmacol Ther* 1983;33:674-687.

18. Hutt PB. Investigations and reports respecting FDA regulation of new drugs (part I). *Clin Pharmacol Ther* 1983;33:537-548.

19. U.S. Food and Drug Administration. Milestones in U.S. food and drug law history. *US. Food and Drug Administration.* 2001. 1.

20. National Cancer Institute. FDA regulations and U.S. drug law: A short history. *National Cancer Institute.* 2001. 1.

21. Marks HM. The Progress of Experiment. Science and Therapeutic Reform in the United States, 1900-1990. Cambridge: Cambridge University Press, 1997.

22. Lasagna L. Congress, the FDA, and new drug development: before and after 1962. *Perspectives in Biology* & *Medicine* 1989;32:322-343.

23. Lasagna L, Frei E. The impact of regulations, tradition, and experimental design on clinical cancer trials: report and recommendations resulting from Washington Cancer Trials Conference. *Am J Clin Oncol* 1996;19:325-329.

ETHICAL CHALLENGES IN COMPLEMENTARY AND ALTERNATIVE MEDICINE CANCER SYMPTOM RESEARCH

Jacquelyn Slomka, PhD*
The University of Texas Health Sciences Center at Houston
School of Public Health

**Corresponding author:*
Jacquelyn Slomka, PhD
The University of Texas Health Sciences Center at Houston
School of Public Health
RAS E-607
1200 Herman Pressler
Houston, Texas 77030
Phone: 713-500-9267
Fax: 713-500-9264
E-mail: jslomka@sph.uth.tmc.edu

ETHICAL CHALLENGES IN COMPLEMENTARY AND ALTERNATIVE MEDICINE CANCER SYMPTOM RESEARCH

Ethical issues in human research, including those specifically related to the ethics of clinical trials, are widely discussed in the medical and nursing literature.[1-5] For the most part, the ethical challenges in CAM research are similar to the ethical issues in any clinical research endeavor. Ethics in human research is not simply the process of obtaining informed consent from participants.[1] Nor can ethics necessarily be equated with law or with an investigator's personal beliefs or clinical experience.[6] Guidelines for ethical research with humans have shaped and are currently shaping the field of research ethics both nationally and internationally. These guidelines include the Nuremberg Code,[7] the World Medical Association's Declaration of Helsinki,[8] the Council for International Organizations of Medical Sciences (CIOMS),[9] and in the U.S., the Belmont Report,[10] and the reports of the National Bioethics Advisory Commission (NBAC).[11] Periodically, interest in research ethics is renewed, often when questionable, unethical, or frankly abusive situations come to light. Most investigators are familiar with well-publicized cases of unethical or questionably ethical research such as the Tuskegee Syphilis Study,[12] the Willowbrook hepatitis studies,[13] and research described by Henry Beecher.[14] More recent cases such as Jesse Gelsinger,[15] the Johns Hopkins lead paint study,[16] and the deaths of healthy volunteers[17,18] demonstrate the need for continued attention to ethics, scientific integrity, and the protection of human subjects in the conduct of health care research.

Many ethical questions stem from societal changes that continue to shape the nature of the research endeavor. In recent years the heavy focus on economics in health care and academia has placed even more emphasis on obtaining research funding and publishing results. An emphasis now is on multi-institutional research rather than smaller scale studies.[19] Funding for non-government sponsored research is dominated by commercial interests that may lead to questionable ethical practices.[20] All of these factors contribute to what appears to be a "moral minefield" of issues in research ethics today that will impinge on the ethical issues in CAM research. This paper will review these issues in the context of research on cancer symptom management at the end of life. The nature of CAM therapies, their use in potentially vulnerable populations such as children and terminally ill cancer patients, the perception of CAM as unorthodox by many patients and professionals, and the impetus to develop CAM therapies as alternatives to more costly orthodox treatments present special ethical challenges for researchers and participants alike. Suggestions for further directions and methods for dealing with ethical issues in CAM research will be delineated.

METHODOLOGY, CAM, AND ETHICS

Two of several criteria for a research study to meet ethical standards are that the study must have value and its design must be appropriate.[1] The randomized clinical trial (RCT) is considered widely as the "gold standard" for assessing the efficacy of therapeutic measures. Likewise, the randomized trial design is also used in non-clinical settings such as community-based or preventive trials.[21] However, many CAM studies are not subjected to this standard. In a review of evidence for the efficacy of CAM therapy in symptom management at the end of life, Pan et al. found 21 studies of CAM use in adults at the end of life, and only 11 of these were RCTs. These authors suggest that the apparent lack of interest in investigating CAM in spite of some

evidence of efficacy is due to funding agencies' lack of interest in both CAM and palliative medicine, which leads to difficulty in carrying out large-scale studies. In addition, methodological issues make it difficult to compare and contrast results of CAM studies. According to the authors, some of these issues involve a) difficulty in using sham controls in CAM research; b) difficulty in blinding both subjects and investigators during the testing of an unconventional procedure; c) a bias against the publication of CAM studies in refereed, indexed journals, leading to investigators' lack of interest in pursuing such studies; and 4) lack of consistency in the composition of herbal and dietary supplements that would permit controlled testing.[22]

Richardson[23] also provides a thorough review of the methodological issues in the use of RCTs in CAM. She suggests that a major difficulty occurs in insisting on using a positivist, reductionist approach to evaluating CAM treatments when such therapies, by virtue of their unorthodox nature, may not be amenable to such analysis. She concludes that decisions regarding the methodological approach and the use or non-use of placebo controls must be derived from the research question rather than from a strict adherence to the use of the RCT.

Reports of CAM efficacy using designs less strict than the RCT (e.g., uncontrolled studies) are beginning to appear in major medical journals, but often serve to discredit a particular CAM therapy.[24,25] (Interestingly, the unwritten rule about not publishing negative studies seems not to apply to CAM research.) Others have pointed out that preliminary studies with designs other than the RCT can provide critical information that can be used to develop a RCT.[23] However, researchers often conduct non-randomized or non-controlled trials to obtain information for clinical use, then fail to take the additional steps to conduct a RCT (personal communication, Dr. Kristine Nelson). The issue of how best to evaluate a CAM therapy is an important ethical question because of the obligation to provide clinical treatments that are effective and safe. However, it is likely that the issue will remain unresolved as long as conventional methods of evaluation are the standard for assessing an unconventional treatment. Paradoxically, if CAM cannot be subjected to the RCT standard, it may never be able to attain conventional status.

CLINICAL EQUIPOISE AND CAM

The fact that, by definition, CAM is alternative and unorthodox presents another morally problematic situation for researchers. Clinical equipoise, the state of the medical community's lack of consensus about the comparative merits of a therapy before testing it, is an ethical requirement of clinical research. Whether or not the individual physician is able to recommend or discourage the patient's participation in a clinical trial of a controversial treatment is irrelevant because the state of equipoise derives from the medical community and not from an individual physician or investigator.[26] In the U.S. today, the orthodox medical community, for the most part, dismisses or discourages most CAM therapy. Although there are exceptions to this generalization, the acceptance of CAM by the U.S. biomedical community has not reached the level of acceptance of, for example, herbal medicine in Germany or East Asian therapies in Japan by biomedical physicians in those countries.

For a variety of political and social reasons, prior to World War II, allopathic medicine assumed dominance in the U.S.[27] and all other non-biomedical forms of treatment were subordinated to

"scientific" medicine. The establishment of the Office of Alternative Medicine by the National Institutes of Health in 1992 provoked much controversy at a time when CAM tended to be dismissed as "placebo effect."[28] At a time when suburban housewives are using alternative therapies,[29] courses on CAM are being taught in many American medical schools,[30] and the Office of Alternative Medicine has become a Center,[31] the biomedical community retains not just skepticism, but negativity towards CAM. For example, as recently as March 2001, the American Pediatric Association published its recommendations on counseling families who use CAM to treat children with chronic illness or disability.[32] Their opinion is that "Many CAM approaches are based on inconsistent or implausible biomedical explanations, and claims of effectiveness rest on anecdotal information and testimonials" (p. 600). Because so little scientific testing of CAM has been done, this is probably an accurate statement. However, the APA goes on to assert that "Alternative therapies may be directly harmful by causing direct toxic effects, compromising adequate nutrition, interrupting beneficial medications or therapies, or postponing biomedical therapies of proven effectiveness" (p. 600). Given the paucity of data on the effects of CAM, the APA's assertion seems somewhat overstated.

While physicians' acceptance of and interest in certain CAM therapies is increasing, the APA's position statement reflects the general negativity of the medical community toward CAM. Part of this negativity stems from the legal climate of health care delivery. Physicians are ethically and legally held to a standard of care, and their recommendations about the efficacy or safety of CAM therapies may lead to malpractice liability.[33] The practice environment thus becomes part of a circular pattern of non-acceptance, leading to lack of research interest, resulting in further non-acceptance of CAM therapies. One ethical problem for research on CAM is that true clinical equipoise will be difficult or impossible if the medical community's opinion remains weighted against CAM. A greater degree of open-mindedness towards CAM by the medical community (and by the legal community?) is necessary to achieve the balance necessary for clinical equipoise. Such equipoise is not impossible and studies do proceed, as in the case of folic acid use to prevent neural tube defects or zinc to treat colds. However, in many current research studies involving CAM, true clinical equipoise may be impossible to attain.

ETHICS AND PLACEBO USE

The issue of placebo use is significant for evaluating the ethics of RCTs because appropriate study design is needed for a project to meet ethical standards.[1] It is commonly perceived that use of a placebo control is essential to clinical trial design. At the same time, comparing a therapy to a placebo control when effective treatment is available generally is considered to be unethical. The issue in CAM symptom research is whether such research can be justified when an effective intervention already is available. The research community remains divided about the ethics of placebo use in clinical trials when effective treatment exists, though some writers are attempting to reconcile differences among the scientific factions by applying ethics criteria on a case-by-case basis.[34]

The withholding of known effective treatment in clinical research was criticized in Beecher's 1966 exposé,[14] but the topic seemed to evoke only minor interest. The serious controversy began in 1994 when Rothman[35] criticized the "continued unethical use of placebos" in the New England Journal of Medicine and listed examples of research studies that used placebos

inappropriately. In 1997, further controversy erupted with publication of a critique of clinical trials of zidovudine in Africa to prevent maternal transmission of HIV to newborn infants. In these trials, some or all subjects were not provided with antiretroviral medication, even though zidovudine was shown to be effective in reducing maternal-infant transmission of HIV in trials in the U.S. and France.[36]

Much of the discussion about the ethics of placebo use centers on trial design and, in doing so, skirts the more fundamental ethical issue—the obligation to do no harm by withholding known effective treatment. Coupled with the ethical premise that the welfare of the research participant should never be subordinated to the goals of science, Rothman's[35] criticism seems on the mark. In addition to valid concerns about trial design in select (but not all) cases, one possible explanation for the continued withholding of effective treatment in research is that pharmaceutical companies have a vested interest in having their newly developed drugs compared with placebo, rather than with known effective drugs, to which their new products may or may not measure up. While the efforts of the industry certainly have resulted in improved, effective products that benefit patients, many drugs developed by pharmaceutical companies are what have been commonly referred to as "me-too" drugs. These are drugs that have been developed not in response to a particular health need, but in response to a competitor's marketing of a similar, promising new drug, in an effort to claim a portion of the market share of the other company's new product.[37]

Why should CAM researchers be concerned about the placebo use controversy? One reason is that design and methodology are often the focal points for criticism of studies of both orthodox medical treatment and CAM. But more importantly, withholding known effective treatment, when it is part of a research trial involving treatment of symptoms in terminally ill subjects, can increase pain and discomfort unnecessarily. The controversy over the drug trials of ondansetron is a case in point. This drug was developed to treat chemotherapy-induced symptoms of nausea and vomiting and was compared against a placebo.[38] The trials were criticized as unethical because known effective treatment was withheld from participants.[39] While some authors argue that placebo use in the ondansetron case was justified because nausea and vomiting are transient, non-life-threatening symptoms, it has also been argued that the physical and psychological effects of withholding treatment may be more severe in terminally ill cancer patients than they would be in healthier subjects. (One author notes that a Pavlovian-type response to chemotherapy can develop if nausea and vomiting are not aggressively prevented.)[39]

It is not clear how well the participants in the ondansetron trials understood the risks and burdens of the trials. The literature suggests that few research subjects are truly informed—the "therapeutic misconception,"[40] by which participants in research believe that an experimental intervention will be of direct benefit to them—is still prevalent. Furthermore, authors who claim that placebos should be used in studies when the burdens caused by withholding effective treatment are mild and benign risk sounding cavalier. It has been noted by critics of withholding effective treatment in research that few patients would choose not to have effective treatment in a trial if given a choice. The assessment of burdens and benefits ethically derives from the patient/subject's viewpoint, not that of the investigator.

The range of CAM therapy is broad—acupuncture, massage, homeopathy, energy therapy, as well as herbal medications are included in the spectrum. Animal studies and Phase I trials of CAM are not frequently seen in the literature, but could assist in answering some questions about therapeutic safety. Ultimately the issue of harm and the use of placebos in CAM trials may indeed have to be considered on a case-by-case basis.

USE OF VULNERABLE PERSONS AS RESEARCH PARTICIPANTS

While all research participants are subject to protection, certain categories of participants may be considered in greater need of protections due to physical, mental, or social factors that could affect the individual's ability to provide informed consent to participate in research. For example, children and the elderly, at opposite ends of the life span, may have physiological characteristics different from middle-aged adults; mentally handicapped persons may not be able to comprehend fully the risks and benefits of research; and economically disadvantaged persons may be more vulnerable to undue influence to participate in research. Children, women, elderly persons, pregnant women, the mentally incompetent, prisoners and other institutionalized persons, and ethnic minorities and economically disadvantaged groups are some of the populations that may fall under the designation of "vulnerable." Vulnerable populations once were excluded from research as a means of protecting them, according to the ethical principle of "do no harm." The current trend, however, is to focus on justice and fairness, and to actively recruit to include members of certain vulnerable populations so they may share in the benefits derived from research.[41]

In CAM and cancer symptom research, two vulnerable populations likely to be recruited for studies are children and the terminally ill. Consideration of the ethical issues concerning the use of children has only recently received attention because in the past, investigators focused their research on populations of white, adult male subjects. Because so little is known about CAM research in children, investigators may need to conduct animal studies, Phase I trials, or related adult or adolescent studies prior to testing a CAM intervention in younger children (personal communication, Dr. Lonnie Zeltzer). However, studies would have to occur within the federal research regulations for children that weigh different levels of risk and benefit.[42] In an effort to provide added protection from research risks, some medical centers allow a child to participate in only one study. Because of this limited access to subjects, it is incumbent upon CAM investigators to demonstrate the value of CAM research as a legitimate adjunct to standard medical research (personal communication, Dr. Patricia McGrath).

Patients who are near the end of life are also vulnerable research subjects because they have severe physical and psychosocial symptoms that influence their participation.[43] While these patients theoretically may be willing to participate in research, symptoms such as fatigue and changes in mental status may present barriers to participation.[43]

In dealing with vulnerable populations, great care must be taken to evaluate the burden-benefit ratio and to ensure informed participation. Patients' interest and openness to CAM therapy may influence their desire to participate. Patients also may experience undue coercion to participate in a study if they have been asked to do so by their caregivers. In publishing research results, ideally investigators should consider describing how vulnerable patients were recruited and

enrolled in the project in order to clarify the consent process and ensure its validity. However, such a standard is not yet in place. In an analysis of research in palliative care populations, 76% of studies using palliative care patients (n = 42) did not disclose how patients were recruited or how informed consent was obtained.[44] In addition, a careful assessment of the likelihood of the research to answer significant research questions should be done as part of the risk-benefit evaluation. The ethical premise that the subject should never be subordinated to the science is even more important in research with vulnerable populations.

Another concern with doing CAM symptom research with terminally ill patients is that the patient's condition may deteriorate during the course of the study. The patient may then be unable to adequately report results of the intervention or to express his or her wish to continue with the study or to withdraw. A general guideline used in research involving mentally impaired subjects is that research on incapacitated patients is appropriate if a) the research focuses on the condition causing the incapacity; or b) the research focuses on a medical or social issue produced by the patient's incapacity.[45] To extrapolate from this guideline, the investigator would have to provide adequate justification as to why the research needed to be continued on the dying patient who has become incapacitated. Is there something in the condition of dying that would justify continuing, or could similar results be obtained using capacitated subjects? An ideal situation would be one in which a CAM symptom research study is applicable to either category of patients—consenting and those unable to consent—and the research can then be conducted on the patients who are able to give consent. For some studies, incapacity itself could be considered the end point of the investigation (personal communication, Dr. Andrea Shelton). Another strategy might be to over-sample terminally ill subjects in order to exclude the use of patients who may lose decisional capacity during the study, if such exclusion is compatible with the research design.

In considering issues of consent for vulnerable populations such as children and the terminally ill, a surrogate or proxy decision maker (usually a parent in the case of children) can be asked to provide informed permission for the patient's participation. The parent or proxy gives "permission," if full "consent" of the patient is not possible. Additionally, it is generally accepted (and may be legally required in some situations) that a child or other vulnerable patient unable to give consent should be asked to give his or her "assent" to the research whenever feasible.

In addition to involving a surrogate or proxy decision maker, two other strategies for protecting subjects who become incapacitated during a study may be relevant. One is the use of advance directives specifically for research that address the issue of continued participation when the subject is unable to consent or withdraw.[45] The other is community-based consent, in which standards for participation are developed through community consensus—a technique that has been suggested in addressing consent in emergency situations.[46] However, none of these strategies rise to the level of having a fully capacitated patient-subject able to provide informed consent to participate in research.

CONFLICTING LOYALTIES

For physicians who assume dual roles of caregiver and investigator, a conflict exists between "doing research" and "doing good" for the patient. Historically, as research moved from small-

scale investigative treatments that directly benefited the patient to large-scale studies in which the individual received no direct benefit—that is, when the focus became "doing good" for society sometimes at the expense of the individual—the ethical concerns in doing research on human subjects increased.[47] The beginning of the acceptance and use of randomized clinical trials (RCTs) after the second World War underscored the conflict between the physicians' duty to act in the patient's best interest and the duty to discover the best treatments for future patients through research.[48]

If a physician has divided loyalties to the patient and to research, the patient may lose trust in the physician's ability to provide the best care. The issue of trust is especially important in research involving terminally ill patients because of their vulnerability. The physician-patient relationship and trust in the health care system can affect participants' acceptance of research and physicians' willingness to refer patients to clinical trials.[49] Lambert et al. demonstrate how public trust in the health care system, in physicians and in the Infantile Paralysis Foundation—together with good marketing—facilitated the polio vaccine trials in 1954, trials that would not meet today's ethical standards for informed consent.[50] Unfortunately, trust in physicians and trust in the health care system has deteriorated. Physicians' and researchers' loyalties are shared with the industries, agencies, and corporations (especially pharmaceutical companies) that fund research.

The economic forces that are currently driving the research endeavor are not likely to benefit CAM research. CAM therapies tend to be inexpensive, low-technology interventions (and therefore are sometimes considered substandard care by patients and providers alike). Corporate funding sources will look for new drugs and technologies that are marketable and profitable—treatments that most likely will exclude CAM. Physician-investigators, subject to pressures to obtain funding, have become increasingly enmeshed in this corporatization of medicine and research.

Physician/investigator and corporate conflict of interest holds the potential for harming the research endeavor if patients perceive physicians as acting from financial motivation rather than for patient benefit. On the other hand, physicians may not be willing to contribute their time and energy to research unless financial incentives are provided. While broad practice guidelines for dealing with conflicts of interest are under development, a current trend in practice is for physicians and investigators to disclose conflicts of interest in order to maintain patient trust in physicians and in the research process.

SUMMARY AND CONCLUSIONS

In addition to the general ethical issues involved in conducting clinical research, the study of CAM for symptom management in palliative care patients presents special ethical challenges due to the vulnerability of the research population and the skepticism and bias of the medical community towards CAM. Six areas of further research are suggested.

The role of researcher conflict of interest and its effect on the physician-investigator-patient-subject relationship: Little is known about how the public views the conflicts of interest inherent in the corporatization of medicine and research. How do pressures to obtain funding and to

publish affect the public's perception of the research endeavor? How do perceptions about conflict of interest specifically affect CAM research?

The effect of professionals' attitudes and perceptions about CAM on the research process: Continued research on CAM will depend to some extent on support of the research process by both patients and physicians. How do the physician's or investigator's biases about CAM influence a patient's choice to participate in CAM research? Do patients and physicians share similar ethical concerns about the use of CAM therapy?

The requirement for clinical equipoise in research: Is equipoise a reality in research today or is it a myth? If public pressure to do CAM research increases, will the concept of clinical equipoise expand to include not only the medical community, but the patient community also?

Protection of vulnerable subjects in CAM/palliative care research: Should investigators be required to publicize their methods of recruitment and process of obtaining informed consent? If a patient becomes decisionally incapacitated during a study, should s/he be automatically withdrawn from the investigation? Should patients/subjects become involved in designing with investigators portions of studies in which they participate (similar to the way public health researchers try to obtain community involvement in research)?

Special issues in the design of CAM trials: Should known effective treatment ever be withheld in the study of symptom relief for dying patients? Are there other "ways of knowing" that can obtain valid and reliable research results without using an RCT? Should "efficacy" be the major criteria for judging CAM or should "effectiveness," in terms of the therapy's meaning for the patient, be considered?

"Controlling for culture" in clinical trials: In the JAMA November 11, 1998 special issue on CAM, several RCTs were described with varying results. Most of these trials involved testing a CAM therapy in a Western (including Australian) population. Only one study was done in the cultural setting of the therapy.[51] More research is needed on the differences of CAM therapies both inside and outside of their cultural context. Acupuncture practiced in the U.S., for example, is stripped of its cultural context. Is acupuncture in China or Japan more effective because implicit societal support is present? How can cross-cultural studies of CAM be designed?

In spite of medical biases, the use of CAM therapies persists. The use of CAM in treating cancer symptoms may hold the prospect of effective, safe treatment for patients. But widespread acceptance of CAM by the biomedical community will not occur without adequate assessment and evaluation through research. Ultimately, the basic reason for CAM research is to enable physicians to assist patients in reaching their health care goals.[52] Much work remains to be done.

The author gratefully acknowledges the other members of the panel on "Expert Opinions on Methodology: Development of CAM Cancer Symptom Research," whose thoughtful comments during discussion of the original version of this paper have aided in its revision. The author is solely responsible for opinions, errors, and omissions.

REFERENCES

1. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA 283:2701-2711, 2000.

2. Lumley J, Bastian H. Competing or complementary? Ethical considerations and the quality of randomized trials. International Journal of Technology Assessment 12:247-263, 1996.

3. Miller FG, Rosenstein DL, DeRenzo EG. Professional integrity in clinical research JAMA 280:1449-1454, 1998.

4. Passamani E. Clinical trials—Are they ethical? NEJM 324:1589-1592, 1991.

5. Hellman S, Hellman DS. Of mice but not men. Problems of the randomized clinical trial. NEJM 324:1585-1589, 1991.

6. Lo B. Resolving Ethical Dilemmas: A Guide for Clinicians. Baltimore, MD: Williams and Wilkens, 1995.

7. http://ohsr.od.nih.gov/nuremberg.php3 (Accessed 7/30/02).

8. http://ohsr.od.nih.gov/helsinki.php3 (Accessed 7/30/02).

9. http://www.cioms.ch/frame_guidelines_july_2002.htm (Accessed 7/30/02).

10. http://ohsr.od.nih.gov/mpa/belmont.php3 (Accessed 7/30/02).

11. http://bioethics.georgetown.edu/nbac/pubs.html (Accessed 7/30/02).

12. Brandt AM. Racism and research. The case of the Tuskegee Syphilis Experiment, in Reverby SM (ed.). Tuskegee's Truths. Rethinking the Tuskegee Syphilis Study. Chapel Hill, NC: The University of North Carolina Press, 2000, pp. 15-33.

13. Krugman S. The Willowbrook hepatitis studies revisited: Ethical aspects. Reviews of Infectious Diseases 8:157-162, 1986.

14. Beecher HK. Ethics and clinical research. NEJM 274:1354-1360, 1966.

15. Gelsinger P. Jesse's intent. Guinea Pig Zero. A Journal for Human Research Subjects, no. 8:7-17, 2000.

16. Glantz LH. Nontherapeutic research with children: *Grimes v Kennedy Krieger Institute*. Am J Public Health 92:1070-1073, 2002.

17. Steinbrook R. Protecting research subjects—the crisis at Johns Hopkins. NEJM 346:716-720, 2002.

18. Steinbrook R. Improving protection for research subjects. NEJM 346:1425-1430, 2002.

19. Moreno J, Caplan AL, Wolpe PR. Members of the Project on Informed Consent, Human Research Ethics Group. Updating protections for human subjects involved in research. JAMA 280:1951-1958, 1998.

20. Chren M, Landefeld CS, Murray TH. Doctors, drug companies and gifts. JAMA 262:3448-3451, 1989.

21. Gordis L. Epidemiology. (2nd ed.) Philadelphia, PA: WB Saunders Company, 2000.

22. Pan CX, Morrison RS, Ness J, Fugh-Berman A, Leipzig RM. Complementary and alternative medicine in the management of pain, dyspnea, and nausea and vomiting near the end of life: A systematic review. Journal of Pain and Symptom Management 20:374-387, 2000.

23. Richardson J. The use of randomized control trials in complementary therapies: exploring the issues. Journal of Advanced Nursing 32:398-406, 2000.

24. Italian Study Group for the Di Bella Multitherapy Trials. Evaluation of an unconventional cancer treatment (the Di Bella multitherapy): results of phase II trials in Italy. BMJ 318:224-228, 1999.

25. Green S. Immunoaugmentative therapy: An unproven cancer treatment. JAMA 270:1719-1723, 1993.

26. Freedman B. Equipoise and the ethics of clinical research. NEJM 317:141-145, 1987.

27. Starr P. The Social Transformation of American Medicine. New York: Basic Books, Inc., 1982, pp. 79-144.

28. Jacobs JJ. Building bridges between two worlds: The NIH's Office of Alternative Medicine. Academic Medicine 70:40-41, 1995.

29. McGuire M. Ritual Healing in Suburban America. New Brunswick, NJ: Rutgers University Press, 1994.

30. Wetzel MS, Eisenberg DM, Kaptchuk TJ. Courses involving complementary and alternative medicine at US medical schools. JAMA 280:784-787, 1998.

31. Marwick C. Alterations are ahead at the OAM. JAMA 280:1553-1554, 1998.

32. American Academy of Pediatrics. Counseling families who choose complementary and alternative medicine for their child with chronic illness or disability. Pediatrics 107:598-601, 2001.

33. Cohen MH, Eisenberg DM. Potential physician malpractice liability associated with complementary and integrative medical therapies. Ann Intern Med 136:596-603, 2002.

34. Emanual EJ, Miller FG. The ethics of placebo-controlled trials—A middle ground. NEJM 345:915-919, 2001.

35. Rothman KJ. The continuing unethical use of placebo controls. NEJM 331:394-398, 1994.

36. Lurie P, Wolfe S. Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries. NEJM 337:853-856, 1997.

37. Public Broadcasting System (PBS). Frontline. Prescription for profit.

38. Beck TM, Ciociola AA, Jones SE, Harvey WH, Tchekmedyian NS, Chang A et al., Ondansetron Study Group. Efficacy of ondansetron in the prevention of emesis in outpatients receiving cyclophosphamide-based chemotherapy. Annals of Internal Medicine 118:407-413, 1993.

39. Markman M. When regulatory requirements conflict with ethical study design: The case of oral ondansetron. Cancer Investigation 12:654-656, 1994.

40. Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: Consent to research and the therapeutic misconception. Hastings Center Report 17:20-24, 1987.

41. Mastroianni A, Kahn J. Swinging on the pendulum. Shifting views of justice in human subjects research. Hastings Center Report 31:21-28, 2001.

42. Sugarman J, Mastroianni AC, Kahn JP (Eds.). Ethics of Research with Human Subjects. Selected Policies and Resources. Frederick, MD: University Publishing Group, Inc., 1998, pp. 174-175.

43. Bruera E. Ethical issues in palliative care research. Journal of Palliative Care 10:7-9, 1994.

44. Kristjanson LJ, Hanson EJ, Balneaves L. Research in palliative care populations: Ethical issues. Journal of Palliative Care 10:10-15, 1994.

45. Dresser R. Mentally-disabled research subjects: The enduring policy issues. JAMA 276:67-72, 1996.

46. Kraybill EN, Bauer BS. Can community consultation substitute for informed consent in emergency medicine research? In: King NMP, Henderson GE, Stein J (eds): Beyond Regulations. Ethics in Human Subjects Research. Chapel Hill, NC: The University of North Carolina Press, 1999, pp. 191-198.

47. Rothman DJ. Ethics and human experimentation. Henry Beecher revisited. NEJM 317:1195-1199, 1987.

48. Shafer A. The ethics of the randomized clinical trial. NEJM 307:719-724, 1982.

49. Taylor KM, Margolese RG, Soskolne CL. Physicians' reasons for not entering eligible patients in a randomized clinical trial of surgery for breast cancer. NEJM 310:1363-1367, 1984.

50. Lambert SM, Markel H. Making history: Thomas Francis, Jr, MD, and the 1954 Salk Poliomyelitis Vaccine Field Trial. Arch Pediatr Adolesc Med 154:512-517, 2000.

51. Cardini F, Weixin H. Moxibustion for correction of breech presentation. A randomized controlled trial. JAMA 280:1580-1584, 1998.

52. Sugarman J, Burk L. Physicians' ethical obligations regarding alternative medicine. JAMA 280:1623-1625, 1998.

STATISTICAL ISSUES IN THE STUDY OF COMPLEMENTARY AND ALTERNATIVE MEDICINE (CAM) INTERVENTIONS

Jeff Sloan, PhD*
Lead Statistician, Cancer Center Statistics Department
Mayo Clinic

Celia C. Kamath, PhD
Mayo Clinic

Paul J. Novotny, MS
Mayo Clinic

Debra L. Barton, PhD
Mayo Clinic

Brent A. Bauer, MD
Mayo Clinic

Pamela Atherton, MS
Mayo Clinic

Michelle Blanch
Mayo Clinic

**Corresponding author:*
Jeff Sloan, PhD
Cancer Center Statistics
Mayo Clinic
200 1st Street SW
Rochester, MN 55905
(507) 284-9985
(507) 266-2477 fax
Email: jsloan@mayo.edu

STATISTICAL ISSUES IN THE STUDY OF COMPLEMENTARY AND ALTERNATIVE MEDICINE (CAM) INTERVENTIONS

ABSTRACT

"There is no such thing as a complementary or alternative p-value."

This paper asserts that the study of complementary and alternative medicine (CAM) requires nothing new in terms of statistical methodology. Existing statistical methods and a sound scientific paradigm are the only prerequisites necessary to undertake high quality, defensible, and credible research into the possible efficacy of these nonstandard treatment regimens. The application and interpretation of statistical methods and results in this literature have been highly variable. The degree of scientific rigor with which these new therapies have been tested has often been inadequate. Study design and statistical methods should take into account the characteristics of the CAM agent under study. A priori assessment and definition of the primary outcome measure is essential. Selection of psychometrically validated measurement tools is an integral part of study design. The research plan should include a predetermined method for assessing the clinical significance of results. Missing data are a particular concern and are best handled by multiple methods in a sensitivity analysis. Specific suggestions for a complete statistical analysis are given. The use of novel methods is discussed briefly by way of a number of illustrative examples from the CAM literature. Finally, an optimistic outlook is predicted for the development of research into CAM therapies.

I. INTRODUCTION

Complementary and alternative medicine (CAM) has become an increasing focus of both patients and the medical community in recent years.[1,2] The use of CAM therapies is particularly widespread with cancer patients and continues to increase although there often has been an overall lack of scientific documentation of their effectiveness.[3-5] Increasingly, the general population is also utilizing and demanding insurance coverage for such therapies.[6] This growing patient demand has created pressure to examine and include CAM therapies into "mainstream" medicine.[6,7] The scientific study of these therapies can also inform clinical practice, where currently few standards of care exist, thus enabling physicians to counsel patients regarding the dose and frequency of interventions such as herbs, massage, or acupressure.

A number of reasons exist for the paucity of scientific confirmation of the efficacy of CAM therapies. In general, standards of care for any therapy derive from efficacy data based on the highest level of verification. Well-designed, randomized studies with adequate sample size and from multiple sites constitute the most rigorous scientific evidence. Correlation, descriptive, and uncontrolled studies, anecdotal reports or case studies are less rigorous and do not lead to definitive conclusions. Research into CAM therapies is characterized by scattered evidence, flawed design, and mixed results, and consists of relatively few randomized studies; a majority of CAM studies utilize retrospective and anecdotal data.[8] This is understandable as the typical practitioner of CAM therapies has not been trained in the methods of scientific enquiry and statistical analysis.

Therefore the hurdles to the development of appropriate guidelines for clinical use of CAM therapies are due neither to the irrelevance of the paradigm of scientific inquiry nor the lack of special statistical methodology. Rather they lie in the classical problems of imprecision of definitions of CAM interventions, the use of subjective measures, and the reality of multiple outcomes. The primary message contained in this paper is that the appropriate statistical methodology for examining the effect of CAM therapies exists in the current paradigm of scientific inquiry and needs to be no different than that used for studying more traditional pharmacological agents or treatment interventions. This paper provides a template for the general scientific process of conducting research on CAM therapies, specifically as it relates to statistical issues. The list of criteria for conducting sound scientific inquiry into CAM research has been provided and discussed elsewhere.[9] Thus, our focus will be on statistical issues. We will present a few challenges in conducting CAM research, along with possible statistical solutions and examples. We will then summarize statistical methodology used in such research. We propose a gold standard for conducting CAM research in terms of study design, statistical analysis, and presentation and interpretation of results, including a discussion on several unique statistical issues, relating to the different phases listed above. We finally present some current examples of exemplary CAM research to demonstrate feasibility of scientific rigor in the conduct of such research.

II. HURDLES IN THE DESIGN AND ANALYSIS OF CAM STUDIES:

The primary hurdle in investigating CAM agents lies in the frequent inability of researchers to delineate exactly what the therapy is, what it is expected to do, and what the mechanism of action

is. Other challenges involved in CAM research are similar to those encountered by research into subjective variables like patients' health related quality of life (HRQOL). However, once the measurement issues and clear delineation of interventions and endpoints have been addressed, the rest lends itself to scientific rigor.[10]

Generally, errors made in the design phase of a study are the most critical, illustrated by a "garbage in–garbage out" scenario. Power considerations and an a priori indication of what results will be deemed a "success" or clinically significant are typically neglected. At the statistical analysis stage, inadvertently missing data and confounding influences are often inappropriately handled. It is important to test the robustness of findings through sensitivity analysis to establish their veracity. Finally, errors at the critical design phase contribute to inadequate interpretation of significant p-values, or of any complex methods used, and hinder the appropriate presentation and interpretation of study results (see Table 1 for a list of challenges typically associated with doing such research).

Table 1. List of Typical Problems Encountered in Doing CAM Research

At the Study Design Phase:

1. Often, the collection and analysis of data precede the appropriate definition of primary and secondary endpoints. Mixed results are difficult to interpret because there is no clarity regarding the primary endpoint.

2. Measurement tools and items do not focus on the identified outcomes because of poor choice of measurement tools.

3. The psychometric properties of new measures or items from existing tools are not established before use.

4. The comparability of study results is often hindered due to the lack of consistency between measurement scales and approaches of the various measurement tools used, necessitating some form of data transformation in order to make comparisons.

5. Often, inappropriate or missing power analysis makes it unclear if the conclusion of the study is a function of mere sample size or a true indication of relationships among dependent and independent factors.

6. Frequently, clinical practice recommendations cannot be derived from study results because the clinical significance of results are not adequately addressed in the design phase. In other words, the magnitude of observed effects cannot be interpreted in terms of clinical significance and indication for changes in clinical practice, since the appropriate effect size was not determined in advance.

7. The important endpoints of any study are often confounded with other variables, making it difficult to ascertain the true impact of intervention on these endpoints. This is because the study

was not adequately stratified to take into account the impact of potentially confounding variables.

At the Statistical Analysis Stage:

8. A large degree of patient attrition and missing data, particularly if missing in a nonrandom fashion, can pose a problem for the application of standard statistical methods.

9. Often, imbalances in size and composition between treatment groups are not taken into account in the statistical analysis.

10. The threshold values that are likely to change the study conclusions are often not explored in terms of sensitivity analysis. This is particularly important for new or emerging therapies.

11. Often, there are not enough data to sustain the number of variables included in a statistical model. For example, most multivariable analytical procedures, such as regression analysis which needs an observations-to-variables ratio of 20:1 in order to generate generalizable results, fall far short of needed sample size.

At the Stage of Presentation and Interpretation of Results:

12. The results of CAM studies often focus on reporting averages, failing to include the distribution and range of possible values of study results. These are often as important as reported mean values, especially for newer therapies like CAM agents.

13. Often, the reporting of confidence intervals and interpretation of statistical significance is neglected, with a focus given to a simple presentation of *p*-values without any interpretation.

14. Typically, the statistical significance of results are emphasized, with no consideration given to any clinical significance of the study findings.

Fortunately, most of these problems can be easily addressed as we demonstrate through several examples below, emphasizing the fact that sound research on CAM therapies is an achievable goal. While it is important to be aware of challenges to such research, it is equally important to recognize that most of these problems are surmountable. At the same time, it is important to remember that it is essential to look at the collective body of evidence gained from multiple statistical approaches in order to collect complete and sufficient evidence to demonstrate the efficacy of a treatment. The next section proposes a feasible approach for conducting CAM studies and approaching potential hurdles that might be encountered at different stages in the process.

III. A PROPOSED MODEL FOR CONDUCTING CAM CLINICAL STUDIES:

This section presents what we consider to be a gold standard for carrying out statistical analysis for CAM clinical studies. CAM research needs to be held to the same standard as any other program of clinical research. These therapies do not warrant any special considerations or major

statistical challenges. Current statistical methodology is sufficiently mature to handle any examination of a CAM therapy with requisite scientific rigor.

The delineation of general guidelines for a uniform approach is in Table 2. A priori linkage among research hypothesis, tool selection, study design, statistical analysis, and presentation of results is of primary importance. In addition to the proposed model, we highlight the special statistical considerations to be taken at each stage of a CAM study.

Table 2. A Proposed Model for Conducting CAM Clinical Studies

At the Study Design Phase:
1. Appropriate a priori linkage must be established between the research hypothesis, chosen endpoints, measurement tools, study design, and statistical analysis.
2. A single primary endpoint and appropriate secondary endpoints must be identified in advance.
3. Study design and planned statistical analysis must be focused on the primary endpoint.
4. Appropriate statistical analyses must be identified for each of the secondary endpoints.
5. Clinically significant effect sizes must be established for primary endpoint.
6. Power specification and sample size must be defined for the primary endpoint.
7. Multiple endpoints must be accounted for by appropriate statistical approach, such as splitting significance levels, if necessary.
8. Potential confounding variables must be identified and accounted for in terms of appropriate stratification.
At the Statistical Analysis Stage:
9. Appropriate statistical tests must be identified for primary and secondary endpoints. The analysis of primary endpoint must be kept separate from supporting analyses.
10. Missing data must be examined for patterns of attrition, specifically, whether data is missing at random or in a systematic fashion. Adjustments for missing data must be made accordingly.
11. The assumptions underlying chosen statistical procedures must be examined for validity.
12. A comprehensive sensitivity analysis using different statistical procedures, with their underlying assumptions must be conducted to establish the robustness of findings.
13. Intent-to-treat analysis should be the statistical approach.

14. Separate analyses on appropriately pre-specified sub-populations (e.g., minority or under-served patients) must be conducted.

At the Stage of Presentation and Interpretation of Results:

15. Descriptive as well as distribution statistics must be presented for primary and secondary variables.

16. Study results must include results pertaining to the "average" and "typical" patient.

17. Graphical displays of results for individual patient must be presented, if possible, to demonstrate the range of effects over study sample.

18. Results from basic statistical analysis must precede the presentation of results from more complex modeling. Discrepancies in findings between basic and complex modeling must be discussed and explained.

19. Sensitivity analysis and threshold analysis must be presented and discussed.

20. The statistical significance of results must include interpretation of *p*-values.

21. The clinical significance of results must be presented and discussed.

Study Design:

1. Research Design and Assessment Strategy: The assessment strategy of CAM agents must be appropriate for the stage of investigation into the therapy. For example, in order to test treatments for which scientific study is in its relative infancy (e.g., massage therapy), it may not be prudent to test the treatment against a placebo or sham therapy. Comparison with a sham therapy, which has been generally agreed as relevant only to demonstrate the mechanism of action of the therapy, comes only after demonstrating that the therapy has some potential to work.

The fact that an agent has been demonstrated to modify an endpoint is sufficient to introduce it into more definitive clinical trials. In the case of massage therapy (MT), after conducting pilot studies, if one wanted to investigate causal links between tactile stimulation and communication with the patient, then a test of this hypothesis could be designed as a randomized study involving a true MT arm, a sham MT arm, and a volunteer visit arm. The differences in efficacy across these three treatments would then allow for the estimation of the effects of the individual treatment components of MT.

2. Inadequate Background Information on Endpoints: The design of CAM studies is often complicated by the lack of background information of the mechanisms of action and toxicity profiles of these agents, which may mandate slightly different trajectories of inquiry. It is fair to say that primary endpoints and designs for CAM studies are likely to change over a series of clinical trials as further knowledge is gained regarding the measurement, distribution, and

appropriate analytical approaches to these studies. For example, in their analysis of the collective experiences with several hot flash activity trials involving oncology patients, Sloan et al. [11] have developed a wealth of information to generate recommendations on design issues, effect sizes, power recommendations, sample sizes, and appropriate methods, based on the type of therapeutic agent to be examined (see Figures 1 and 2).

Figure 1. Median Hotflash Score for Placebo Controlled Trials Of therapeutic Agents to Control Hot Flashes.

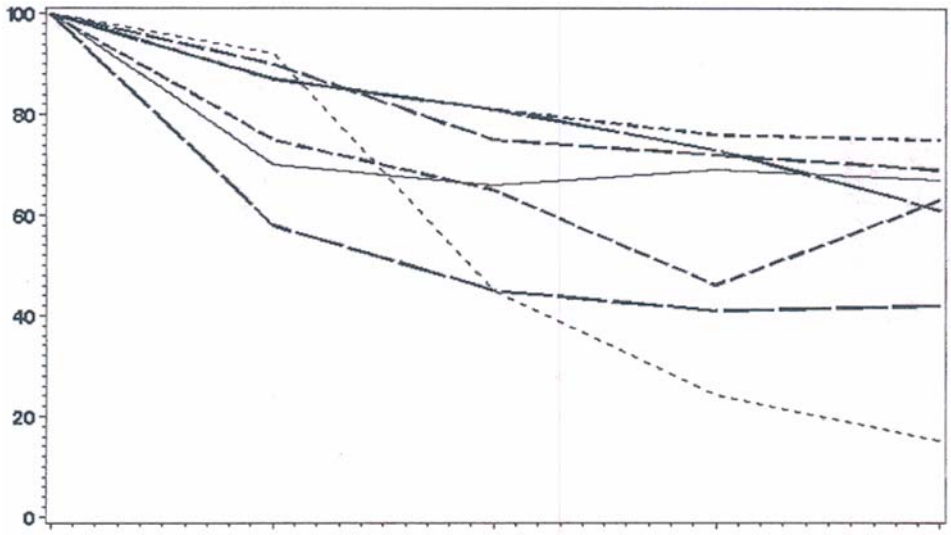
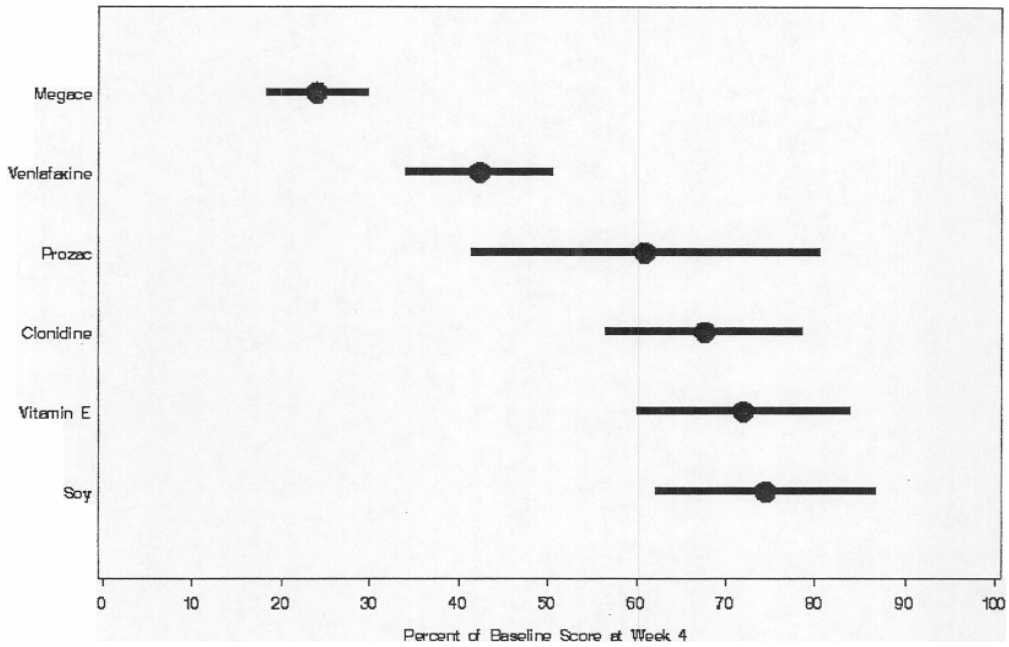


Figure 2. Forrest Plot of Percent of Baseline Score at Week 4 with Corresponding 95% Confidence Limits of Placebo-Controlled Trials Agents to Control Hot Flashes.

Treatment



In general, relatively nebulous CAM agents, like some herbal preparations, will follow the same established trajectory of investigation as other new pharmaceutical agents: studies to examine the mechanism of action (pre-clinical studies), safety (phase 1), rate of response (phase II), and efficacy (phase III). Other agents, like St. John's wort, are sufficiently mature to be tested using phase III randomized clinical trials, since sufficient scientific evidence has been collected to demonstrate that the agent is safe (when not used in conjunction with pharmaceutical drugs) and holds sufficient promise of efficacy.[12,13] Researchers generally agree that the design of clinical trials for certain CAM interventions such as energy work are problematic in terms of the choice of comparator arm for "placebo" or "control group," because of the relative newness of the therapy. "Sham" or "mock therapy" and the current standard of care appear to be the best current options for use as the control group, until further knowledge is accumulated regarding these therapies.

3. Identification of Appropriate Endpoints: The identification and definition of CAM study endpoints are challenging for study design relative to more standard treatments. Many CAM agents have non-specific methods of action and undefined outcomes, making the choice of appropriate endpoints challenging. For example, studies on MT, which has been demonstrated to be associated with a general sense of well-being, are challenged because of the lack of clarity regarding specific aspects of well being impacted by massage therapy. Potential targets of research into MT have included a multitude of measures including pain, anxiety, depression, blood pressure, EEG deviations from normal, salivary cortisol, oxygen saturation, general quality of life, symptom control, and attitude measures.[14-22] The need for preliminary research on these endpoints is underscored here, in order to identify a smaller subset of promising targets impacted by MT that could be clinically investigated further.

CAM study endpoints are often multidimensional, making their definition and choice difficult. Solutions to this problem include the use of bivariate or multivariate endpoints,[23] or the choice of one endpoint as primary with the others as supportive evidence.[24] The methods to do this are described below.

4. Secondary Endpoints: Secondary endpoints, often numerous, are an important consideration in CAM research as much of it is hypothesis generating in nature. It is important for studies at this stage, to consider all potential endpoints in order to obtain a complete profile of efficacy of the therapeutic agent. A useful method to delineate the set of potential endpoints for new agents is the use of focus groups. However, pilot studies must investigate patient burden since numerous endpoints increase the number of measurement tools. Once identified, a good rule of thumb is to include only those endpoints that demonstrate promising strength of effect, within the constraints of a realistic sample size.

5. Quality of Life (QOL) Endpoints: CAM agents usually impact patients' QOL making this a popular endpoint for such studies. QOL is particularly useful for studying the efficacy of CAM agents in cancer patients,[2,7,25] given the psychological ramifications and morbidity related to cancer diagnosis and treatment affecting the health-related quality of these patients. QOL has also been demonstrated to be prognostic for patient survival,[6,27] which is important to cancer patients, given the urgent nature of their prognosis.[8] Patients' overall well-being can be captured by the use of simple QOL measurement tools.[10,28,29] A useful primer for clinicians on this topic

has recently been published.[30] QOL is, however, a multi-faceted and multivariate entity comprising numerous constructs.[31] There remains a substantial amount of controversy regarding the accuracy and clinical significance of measuring QOL.[32,33]

Several issues must be considered in choosing QOL endpoints for CAM studies.[34-36] Typically, one must a priori identify the elements of QOL likely to be of interest and affected by the treatment under consideration. Choice of items from disease-specific tools or existing generic tools should be relevant to the dimensions under study. Irrelevant constructs increase the likelihood of observing no impact because items of zero variance make the results less sensitive to change and bias the study towards an inflated type II error rate.[27,37]

A "menu" approach to compiling appropriate QOL measures is useful. In general we suggest measures that are simple, brief, and targeted at the global constructs of health-related QOL that are likely to change, without undue burden to patients.[27,36,37] This entails describing the specific constructs of interest that are likely to change, compiling items from existing QOL instruments if possible, while ensuring the integrity of the psychometric properties of the original tool. Custom-designed items for constructs not covered by existing instruments, should be done with due psychometric diligence. A uniform response measurement scale should be used for consistency and ease of use.

6. Clinically Significant Effects: An essential initial step in CAM research is to identify how much of an effect will be considered sufficient evidence of efficacy in advance of the trial. A series of papers published in the *Mayo Clinic Proceedings*[38-44] details the various aspects of assessing clinical significance. Several authors have tried to define clinical significance. Some have defined it in terms of anchoring the endpoint of interest to an observable clinical outcome.[45,46] Others have used a statistical approach.[27,36,37,47] The clinical literature also provides guidance for this determination in select areas in the form of expert opinion. A comparison of these methods used in cancer outcome studies indicates that all approaches give similar answers.[48] While such definitions are helpful for academic purposes, they leave much to be desired for facilitating clinical applications.

By combining the empirical rule of statistical theory with the classification of effect sizes due to Cohen,[49] Sloan (unpublished) has proposed a unifying theory, which proposes that a half-standard deviation is a reasonable ballpark estimate for a "duck," signifying a moderate clinically significant effect (compared to a "worm," signifying small effects and an "elephant," signifying large effects). This is especially relevant for CAM research where the constructs under study may not be as well defined as in other clinical research settings. Modifications to the foundational half standard deviation definition of a "duck" may be warranted for specific situations. A sound analytical process needs to be followed so that the modifications retain their statistical veracity, and are not the result of mere guesswork.

Validation for the "worm-duck-elephant" approach to establishing clinical significance lies in the fact that, even after accounting for sources of measurement error, the magnitude of suggested effect size estimates are robust. When no further information is available for a given tool in a population or clinical context, as is the case when investigating novel CAM therapies, this forms a simple, practical approach to identify a suitable effect size for CAM studies.

7. Power Considerations: The power of any study is closely related to the chosen effect sizes and associated variance estimates. Empirical estimates available from the literature or pilot studies should form the basis for setting estimates for a particular study. However, even these preliminary estimates of variance/standard deviation are suspect in terms of the degree to which they are reliable, valid, and applicable to the proposed study. A. Vickers (personal communication) points out that estimates of standard deviation, drawn from pilot studies, result in more than a 50% probability of an under-powered study. Gould[50] suggests the use of interim analyses to accommodate inaccurate pilot data estimates, pointing out similar dangers. Thus pilot data may not always produce the best estimates for power calculations.

In the absence of such estimates, the worm, duck, and elephant framework can provide a guideline for effect-size estimates to use in study design and power analysis. This approach has been validated with an extensive search of the literature involving over 50 empirical studies.[51] Irrespective of the power approach used to determine clinical significance, the most important step is to decide up front the benchmarks that will be considered evidence that the primary endpoint has been significantly altered.

8. Stratification: Stratification of a randomized study by potentially confounding variables is particularly important to CAM studies. This is due to the inherent measurement error associated with endpoints and other problems such as patients' history and natural changes over time. Stratification is complicated by the inability to obtain pre-cancer measurements for many endpoints including confounding variables. These issues can become particularly problematic if an effect size is not expected to be very large or if the implementation of an intervention depends greatly on the patient's performance or abilities.

There is usually a limit to the number of feasible stratification factors for a given study. Typically, one half of the number of observations per treatment group (i.e., n/2*)* is the maximum feasible number of such factors.[52] Treatment assignment to allow for stratification is often carried out by a method of randomly permuted incomplete blocks or by using a dynamic allocation procedure which balances the marginal distributions of the stratification factors between the two treatment-sequence groups.[53]

Stratification factors are often not directly measurable, sometimes necessitating the use of proxy measures. For example, illness severity or survival, important stratification factors for cancer outcome studies are often difficult to measure. We have demonstrated that physicians' estimates of expected survival times for patients in their care are good proxies for patient survival.[54] The use of proxy variables should be preceded by investigation into the degree of association with the concerned stratification factor.

Analytical Approaches:

There is, at present, no generally accepted or optimal statistical approach for CAM Data.[27,36,37] Analysis of the data for CAM studies is best carried out in a number of complementary ways in the form of a sensitivity analysis. Replication of results across a number of statistical approaches

increases confidence that the findings are not a function of the assumptions underlying the statistical procedures.

The importance of considering the use of simple straightforward procedures before resorting to more complicated analyses, is underscored. Basic statistical methods such as descriptive statistics and simple hypothesis tests are fundamental to CAM research, even in the presence of more complex statistical methodology. The analysis of the trials examining St. John's wort, massage therapy, and acupuncture provide excellent examples of how basic statistical methods can be applied to such studies.[13,14,55] There is however, a special role for complex statistical methodology in the analysis of CAM data, typically driven by the complexities of research that cannot be conducted in controlled, randomized settings.

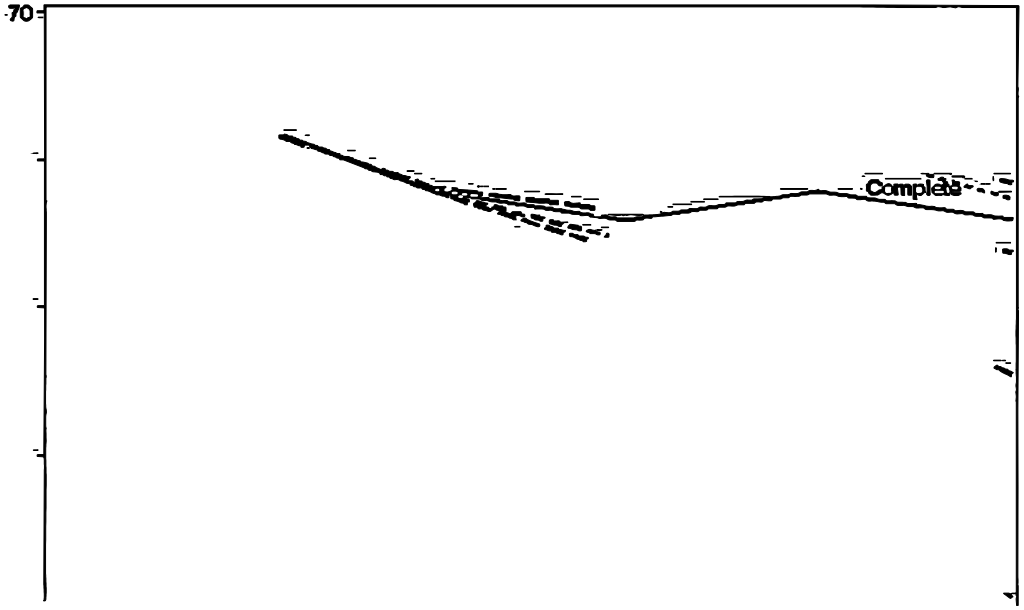
By and large, results from complex or basic analyses should be similar. In other words, we should derive similar estimates for endpoints, whether or not single or multiple potential confounding covariates are included, assuming that the sample size is sufficiently large to accommodate these covariates. If this is not so, the problem is likely to be in the assumptions underlying the modeling of the complex process rather than a real change in the treatment effect. It could also indicate a flaw in the study design, such as neglecting to take into account strong interaction effects. Our suggestion that analyses of the data for outcome research studies be carried out in a number of complementary ways in the form of a sensitivity analysis applies here as well. If the results replicate across a number of statistical approaches, in this instance, basic and complex, then one can have confidence that the models have been correctly formulated and the findings are not merely a function of the assumptions underlying the statistical procedures.

It is critically important to involve an experienced statistical consultant early in the development process in order to deal with complex design and methodology issues. This allows for the necessary statistical considerations such as validation of important underlying assumptions and needed flexibility in analysis to be designed into the study from the onset. Issues relevant to CAM studies that arise at the stage of statistical analysis are discussed below.

9. Missing Data: Missing data is a problem particularly relevant to CAM studies, which experience high patient attrition. There are a number of ways to handle this eventuality in the design and analysis phase of the trial.[56,57] We present a couple of examples of how simple analytical methods can be applied.

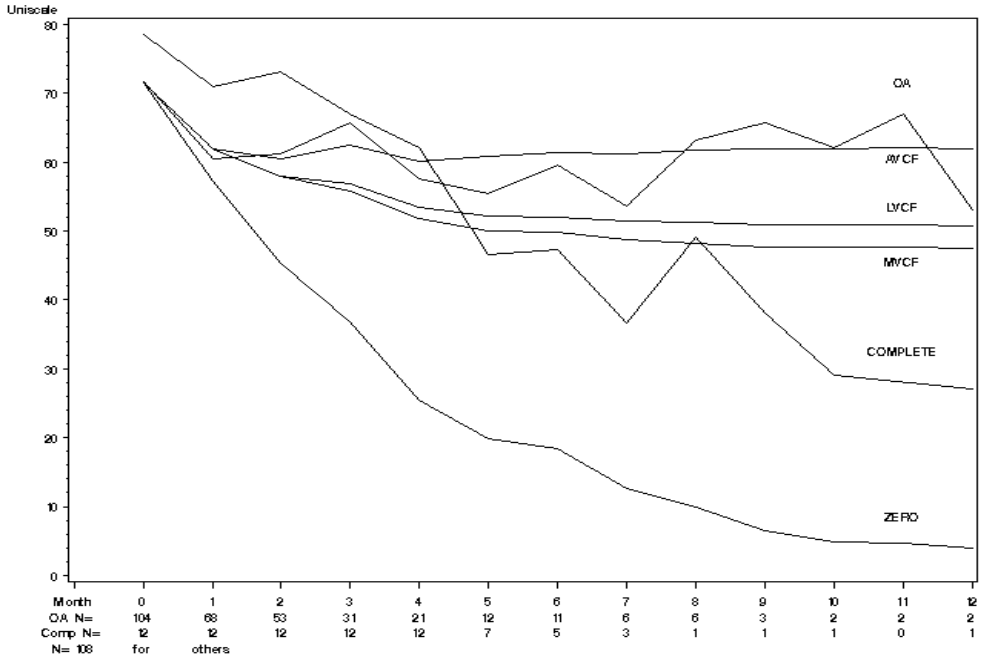
In studies involving seriously ill patients, it is reasonable to assume that missing data occurs because patients are too ill to participate further in the trial. It is important to ascertain if missing data is due to random or systematic forces. Based on this, one can choose to use only available data (OA) or impute missing data by carrying forward the last value (LVCF), the maximum or minimum value (MVCF), the average value (AVCF) obtained for each patient, a zero value (ZERO) after death, or data only from subjects who completed the trial (COMPLETE) to reflect the fact that the patient is no longer living. An example of this is provided in Figure 3, using data from an unpublished clinical trial comparing epoetin alpha with placebo on advanced cancer patients. Each imputation method has different implications for efficacy of treatment. In general, unless more than 20% of the data are missing, the results one gets from any imputation approach will likely be the same.[56]

Figure 3. Mean QOL by Treatment Cycle for Patients Receiving Epoetin Alfa or Placebo Over Four Months.



Differences in imputations also serve as a form of sensitivity analysis. For example, Figure 4 displays average QOL scores for patents with advanced cancer on an (unpublished) clinical trial comparing epoetin alfa to placebo. Sample size in the treatment group went from 322 patients at baseline to 182 patients after four cycles of treatment. We could assume that patients with missing data either have QOL no higher than their last provided observation (line marked LVCF) or that QOL is close to zero (line marked ZERO). Imputing data using both assumptions provides a sensitivity analysis that reflects the best and worst case scenarios, providing markedly different QOL profiles over time. As Figure 4 demonstrates, irrespective of the use of the OA or zero imputation method, we can conclude that QOL is the same for placebo (PL) and epoetin alfa groups (epo).

Figure 4. Mean QOL by Treatment Cycle and Arm Using Different Imputation Procedures (OA and ZERO) for Patients Receiving Epoetin Alfa or Placebo for Four Months.



10. Intent-To-Treat Analysis: There are several advantages and disadvantages of applying the intent to treat (ITT) principle.[58] In brief, ITT characterizes results as a treatment failure if the patient does not report a documented success. Although conservative, this approach provides a rationale for addressing the hard reality of the problem of patient attrition in clinical research.[59] For CAM therapies, this aspect is particularly important because of the varied and potentially unorthodox characteristics of some of the treatments. A pilot study of St. John's wort for hot flashes in cancer patients provides an example of an ITT analysis for cancer patients. This (unpublished) study was originally designed to accrue 10 patients to determine if St. John's wort had any efficacious effect on hot flashes.

If mean hot flashes could be reduced by about 50%, this treatment would be considered promising and would be the subject of future study. After six patients were put on the study, three patients finished the study and had greater than 50% reduction in hot flashes. However, two patients had to stop early because of increasing hot flashes and another finished the study with increased hot flashes. While the mean reduction after four weeks of treatment was greater than 50%, there were really only 3 successes out of the 6 patients. The intent-to-treat analysis thus appropriately reflected findings that St. John's wort failed to produce efficacious results.

11. Analysis of Multiple Endpoints: Several standard multivariate procedures to address the analysis of multiple endpoints exist.[60] At a basic level, decisions must be made on how to split alpha levels (the type I error rate) while interpreting the results. A popular method involves

dividing the significance level by the number of hypothesis tests performed.[61] Thus three primary endpoints would warrant a split of the overall experiment-wise alpha of 5% into three tests using a comparison-wise significance level of 5/3 = 1.67% each. This implies that one should ignore observed *p*-values between 0.05 and 0.0167. However, clinical investigators find it hard to ignore *p*-values less than 2% even in the context of multiple hypotheses tests, choosing to suggest that such *p*-values indicate a "trend towards significance." Such results should be more appropriately discussed in terms of observed effect size and related clinical significance. Establishing an a priori clinically significant effect size makes the splitting of the *p*-value redundant.

A useful approach in analyzing multiple endpoints is to analyze each using univariate procedures, later supplementing results with multivariate analyses. O'Brien proposed a general method for dealing with the analysis of multiple endpoints in clinical trials,[62] which allows for a single *p*-value to express the degree of difference in the relative rankings of observations from treatment groups across an array of endpoints. This technique can be found in a study of venflaxine for hot flash activity in postmenopausal women,[63] where three primary measures of efficacy, (hot flash frequency, severity, and hot flash scores) were examined. Secondary endpoints included patient QOL and toxicity. Table 3 demonstrates statistical significance was achieved in terms of the O'Brien *p*-value for the composite of the first three endpoints. The addition of QOL to the composite reduced the *p*-value. The addition of toxicity variables reduced the *p*-value further, to non-significance. Small differences in toxicity between the treatments overpowered the impressively significant results for the other endpoints of the study.

Table 3. O'Brien *p*-values for Hot Flash Example

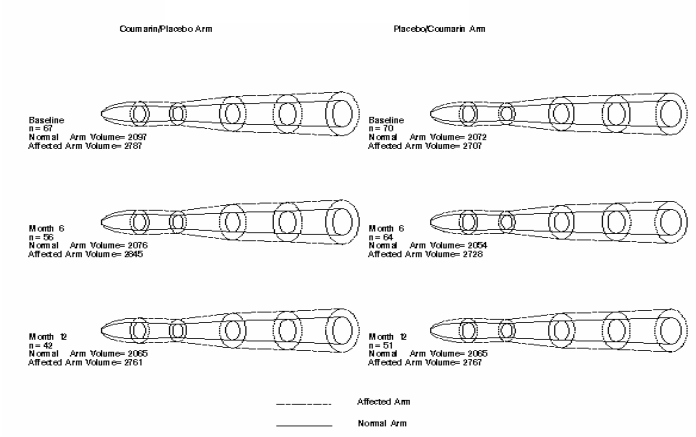
Endpoints Included
Hot flash frequency
Hot flash severity
Hot flash score
Uniscale QOL
Hot flash effect on QOL
Toxicity incidence

The use of novel methods to deal with multiple endpoints is illustrated with this example involving postmastectomy breast cancer patients who often experience periods of lymphedema and augmented hot flash activity. Lymphedema manifests itself in many ways. Table 4 indicates the list of endpoints used in lymphedema studies.[64] A geometrical combination of these values combined into a single endpoint is displayed graphically in Figure 5, which easily replaces individual *p*-values for each of the endpoints.[64] The images displayed are the median values for the swollen limbs superimposed on data for the unaffected arms for patients at baseline and at six and twelve months post treatment in a classic two-period crossover design. As evident in the six images portrayed, there are no differences in baseline and post-treatment images, indicating a lack of treatment efficacy. Other modeling procedures to deal with multiple endpoints are more sophisticated. They explore causal relationships between variables, using structural equation, factor, and path models for the examination of latent variables.[65]

Table 4. Possible Clinical Endpoints for Lymphedema

Circumference of Affected Hand
Circumference of Affected Wrist
Circumference of Affected Arm at 30 cm
Circumference of Affected Arm at 40 cm
Circumference of Affected Arm at 50 cm
Ratio of Circumference of Affected to Normal Hand
Ratio of Circumference of Affected to Normal Wrist
Ratio of Circumference of Affected to Normal Arm at 30 cm
Ratio of Circumference of Affected to Normal Arm at 40 cm
Ratio of Circumference of Affected to Normal Arm at 50 cm
Distal Edema: ratio of the sum of the circumferences at hand, wrist, and 30 cm divided by the corresponding sum of the normal arm
Total Edema: ratio of the sum of all circumferences on the affected arm divided by the sum on the normal arm
Volume of Affected Arm estimated from the circumference measurements
Volume of Affected Arm Divided by the Volume of Normal Arm
Patient Reported pressure pain in arm, heaviness in arm, arm tightness, loss of arm mobility, arm swelling
Patient Preference of which crossover period they preferred
Patient Rating of whether they felt the tablets were helping

Figure 5. Arm Medians Coumarin/Placebo Arm.

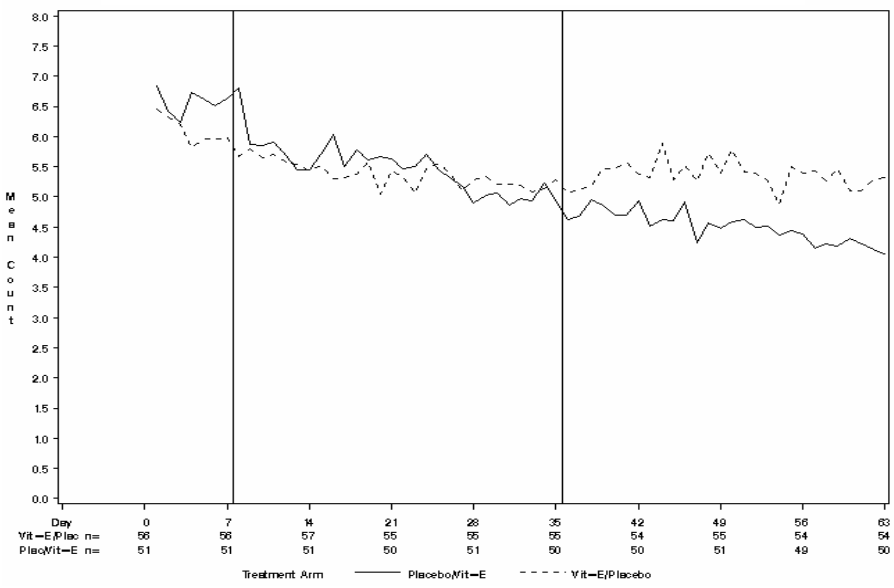


Presentation and Interpretation of Results:

The presentation and interpretation of results are as important as any other stage of a research study in order to establish the utility and relevance of CAM research. The following issues relate to the presentation and interpretation of CAM study results:

12. Clinical Significance of Results: The challenge associated with interpretation of effect sizes of any study is the communication of how these estimates can be incorporated into clinical research and practice. Statistical significance of results must be translated into clinical significance. Using the example of hot flash activity discussed earlier, we know that placebo can result in an average reduction of one hot flash per day and a 25% and 30% reduction in hot flash score. A study to examine efficacy of vitamin E in reducing hot flashes was powered to detect differences of at least this magnitude.[66-68] We observed that women reported a reduction from roughly six hot flashes per day to around 4.7 hot flashes through the use of a placebo (Figure 6). The results were on the cusp of statistical significance, producing *p*-values of between 0.055 and 0.045, depending upon the statistical procedure utilized. However, more important than statistical significance was the fact that a reduction of 0.7 hot flashes per day was clinically important, thus resulting in recommendation of vitamin E on the grounds of clinical rather than statistical significance.

Figure 6. Mean Total Hot Flashes by Week for Patients on Vitamin E to Reduce Hot Flashes Over 9 Weeks.



By and large, translating numerical estimates of clinical significance into simple statements expressed in values of the original measurement tool is vital to make the method accessible to clinicians. In terms of the "duck-worm-elephant" approach used earlier, we need to relate the "duck" as closely to the experimental setting as possible, so that we will recognize it if we see it. This is particularly important for CAM therapies, which are in their relative infancy.

13. Graphical Presentation: One of the most under-utilized methods for presenting the results of clinical trials is graphical presentation. Simple graphics displaying data for individual patient experiences are often more convincing and illustrative than plots of average values for groups. This is particularly appropriate for CAM studies, as the number of individuals involved is typically small. One example is the Event Chart, used to derive a preliminary understanding of data and to track the timing of significant events for individual patients (Figure 7), which is not possible with the traditional scatter plots typically used with preliminary data. A second example is the mirror image stream plot, illustrated in Figure 8 via longitudinal data for two pilot studies of hot flash treatments. The graph displays individual patient hot flash scores as a percentage of baseline scores over time for one of the treatment arms (left of the vertical axis) versus the other treatment arm (right of vertical axis). The horizontal reference line indicates stable hot flashes. If the two treatment groups were equivalent, we should see a symmetric display on either side of the vertical reference line. The variability reflected in the picture also indicates that the efficacy of these two treatments are both promising but the favorable effects are not seen in all patients.

Figure 7. EORTC Changes for Colorectal Patients.

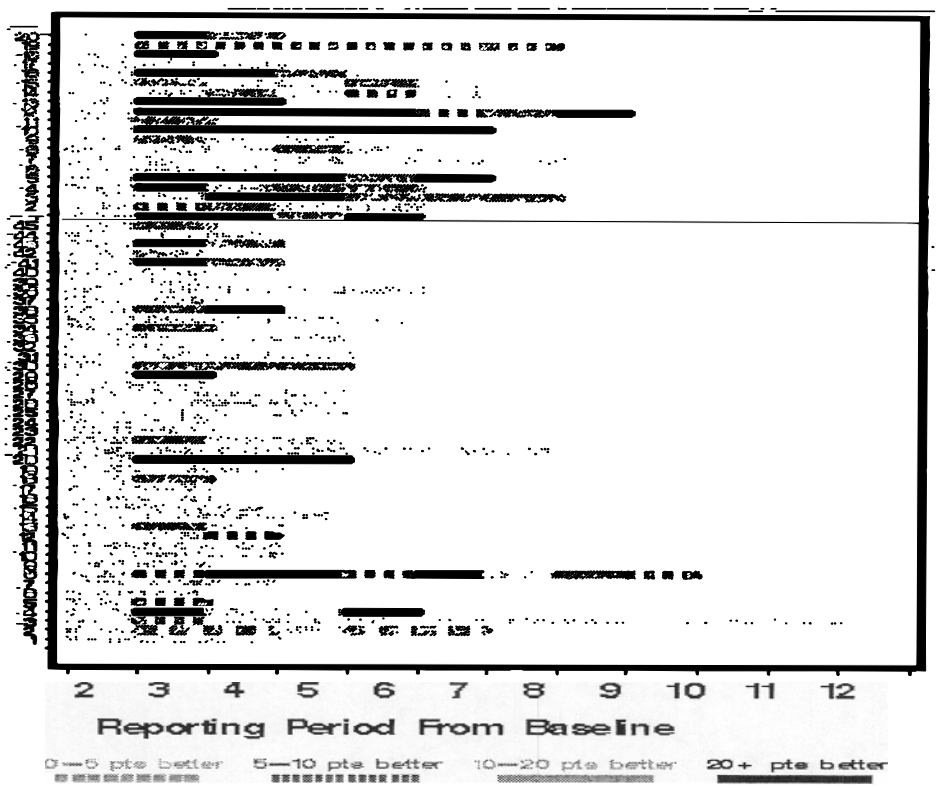
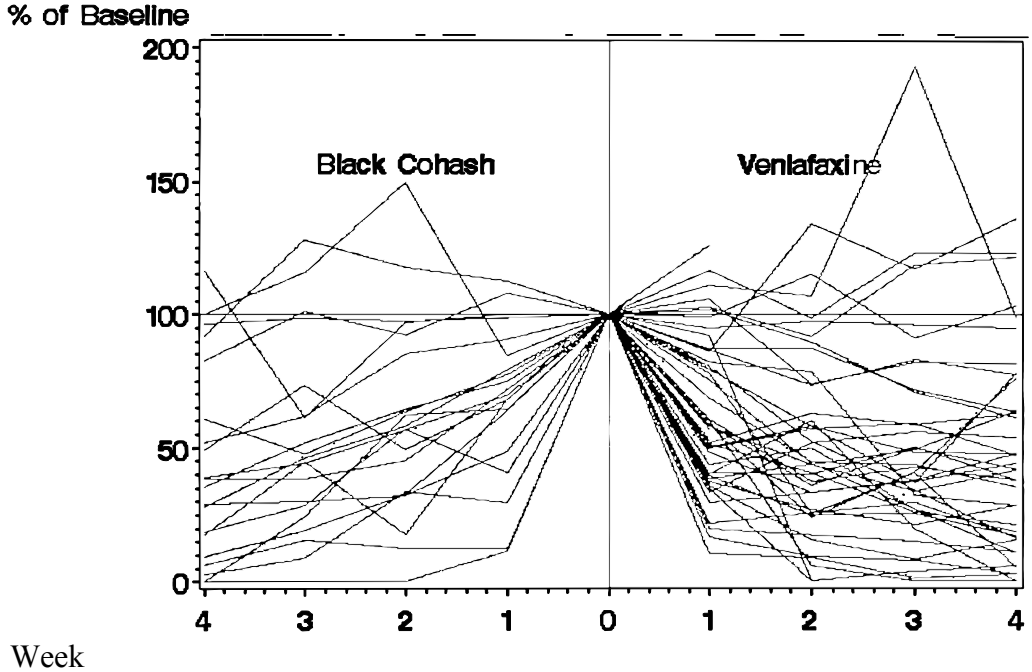


Figure 8. Black Cohash vs. Venlafaxine as Therapeutic Agents For Treating Patients for Hot Flashes. Bug Plot of Change from Baseline.



IV. CURRENT STATUS OF DESIGN AND STATISTICAL APPROACHES TO CAM RESEARCH:

Having discussed the challenges and appropriate statistical approaches to conducting CAM research, we present selective exemplary work done in the area to date. While it is true that relatively few CAM studies have been conducted with appropriate scientific rigor,[69] examples of well-conducted studies demonstrate that such research can be done. In a systematic review of over 5,000 CAM clinical trials over the period from 1966 to 1998, Bloom found that most studies were not randomized, controlled or blinded. He concluded that all but 250 trials failed to meet even minimal standards for scientific enquiry. Nonetheless the literature summarizing recent advances in CAM research[70,71] is encouraging, with recently published works meeting higher standards of scientific enquiry. Three particular examples, worthy of merit are discussed below.

The first example concerns a series of studies examining the role of St. John's wort for depression,[12,13,72,73] which followed numerous investigations into the mechanism of action, the pharmacokinetics, and safety issues concerning this drug. Collectively, the studies met the same scientific rigor as that of any "mainstream" clinical trial, utilizing randomization, placebo controls, intent-to-treat analysis, definition of responses, valid and reliable measures and standard statistical analysis. The collective body of evidence suggests that the effects of St. John's wort, for the treatment of clinical depression, is more modest than earlier believed, better defining its place in treatment with mild depression or mood alteration than severe clinical depression. The importance of this series of trials is that it reflects one of the first times that a

CAM therapy has gone through a review process that is similar to that used for traditional pharmaceutical agents.

The second example involves the studies by Field,[16,74,75] on the use of massage therapy (MT) in neonates. Again we see studies conducted with appropriate scientific rigor, using methodology involving randomization, sham therapy or mock control groups, blinding, and the use of valid and reliable measures. Field demonstrated that MT administered to critically ill pre-term infants results in a clinically significant weight gain compared to placebo or sham therapy. These successful results have led to research on the use of massage therapy in additional clinical populations, such as bone marrow transplant (BMT) patients.[14] This randomized controlled trial (RCT) demonstrated that massage therapy reduced pain and anxiety in BMT patients. Collectively, these studies demonstrate that proponents of CAM therapies are capable of being carried out via the rubric of the controlled clinical trial.

A final example involves an assessment of acupuncture and amitriptyline for peripheral neuropathy in HIV patients.[55] A complex two-factor linear model with repeated measures, comparing treatment with a sham treatment and placebo, was employed in a manner consistent with that used for conventional therapies. The results of analysis of variance modeling of scores, drawn from validated instrumentation, concluded that acupuncture did not significantly reduce pain in these patients compared to amitriptyline or placebo. This example makes a direct comparison between a "traditional" pharmaceutical treatment and a CAM therapy and demonstrates that designing a study with appropriate statistical methods, in and of itself, does not guarantee results. This study could be criticized for methodological flaws related to lack of preliminary data and standardization of treatment (participants may have had access to over-the-counter amitriptyline, and pain medication changes were not noted and/or controlled), but these criticisms did not relate to statistical methods used in the study.

These examples and the proposed statistical model suggested above provide ample evidence that the statistical analysis of CAM studies can and should be held to the same standard as research into traditional therapeutic agents

V. CONCLUSIONS

This paper has attempted to show the current status of statistical issues within CAM research, including existing inadequacies and the reasons such research has led to more confusion and conflict than utility. The difficulties encountered appear to be in two arenas, namely the translation from research question to study design, measurement tools and the consequent implications for the appropriate statistical analysis, as well as communication of results and their interpretability. Fortunately most of these problems can be easily addressed as is evidenced in several examples provided in this paper.

This paper also presents a model for approaching CAM research, given the practical constraints of conducting clinical investigations. The future for CAM research is optimistic given the present level of interest in using these therapies and the feasibility of conducting scientifically rigorous studies on these therapies. Application of established scientific principles is all that is needed to broach the gap between CAM practitioners and research consumers.

REFERENCES

1. Ernst E. Research into complementary/alternative medicine: an attempt to dispel the myths. Int J Clin Pract 55:376-379, 2001.

2. Harris P, Rees R. The prevalence of complementary and alternative medicine use among the general population: a systematic review of the literature. Complement Ther Med 8:88-96, 2000.

3. Downer SM, Cody MM, McCluskey P, et al. Pursuit and practice of complementary therapies by cancer patients receiving conventional treatment. BMJ 309:86-89, 1994.

4. Eisenberg D, Foster D. The role of complementary of alternative medicine, Virtual Annual Session, 1999.

5. Paramore LC. Use of alternative therapies: estimates from the 1994 Robert Wood Johnson Foundation National Access to Care Survey. J Pain Symptom Manage 13:83-89, 1997.

6. Pelletier KR, Astin JA, Haskell WL. Current trends in the integration and reimbursement of complementary and alternative medicine by managed care organizations (MCOs) and insurance providers: 1998 update and cohort analysis. Am J Health Promot 14:125-133, 1999.

7. Meenan R. Developing appropriate measures of the benefits of complementary and alternative medicine. J Health Serv Res Policy 6:38-43, 2001.

8. Richardson MA, Sanders T, Palmer JL, et al. Complementary/alternative medicine use in a comprehensive cancer center and the implications for oncology. J Clin Oncol 18:2505-2514, 2000.

9. Beyerstein BL. Alternative medicine and common errors of reasoning. Acad Med 76:230-237, 2001.

10. Cella DF. Quality of life outcomes: measurement and validation. Oncology (Huntingt) 10:233-246, 1996.

11. Sloan JA, Vargas-Chanes D, Kamath CC, et al. A simple approach for defining clinically relevant effects in quality of life measures (a treatise regarding worms, ducks, and elephants). Unpublished, 2002.

12. Gaster B, Holroyd J. St. John's Wort for depression: a systematic review. Archives of Internal Medicine 160:152-156, 2000.

13. Sheltoin RC, Keller MB, Gelenberg A, Hirshfield R, Thase ME, Russell J, Lydiard RB, Crits-Cristoph P, Gallow R, Todd L, Hellerstein D, Goodnick P, Keitner G, Stahl SM,

Halbreich U. Effectiveness of St. John's wort in major depression: a randomized controlled trial. JAMA 285:1978-1986, 2001.

14. Ahles TA, Tope DM, Pinkson B, et al. Massage therapy for patients undergoing autologous bone marrow transplantation. J Pain Symptom Manage 18:157-163, 1999.

15. Bauer WC, Dracup KA. Physiologic effects of back massage in patients with acute myocardial infarction. Focus Crit Care 14:42-46, 1987.

16. Field T, Ironson G, Scafidi F, et al. Massage therapy reduces anxiety and enhances EEG pattern of alertness and math computations. Int J Neurosci 86:197-205, 1996.

17. Fishman E, Turkheimer E, DeGood DE. Touch relieves stress and pain. J Behav Med 18:69-79, 1995.

18. Lewis P, Nichols E, Mackey G, et al. The effect of turning and backrub on mixed venous oxygen saturation in critically ill patients. Am J Crit Care 6:132-140, 1997.

19. Tyler DO, Winslow EH, Clark AP, et al. Effects of a 1-minute back rub on mixed venous oxygen saturation and heart rate in critically ill patients. Heart Lung 19:562-565, 1990.

20. Weintraub MI. Shiatsu, swedish muscle massage, and trigger point suppression in spinal pain syndrome. American Journal of Pain Management 2:74-78, 1992.

21. Wikkingson S. Aromatherapy and Massage in Palliative Care. Int J of Palliative Nursing 1, 1995.

22. Zanolla R, Monzeglio C, Balzarini A, et al. Evaluation of the results of three different methods of postmastectomy lymphedema treatment. J Surg Oncol 26:210-213, 1984.

23. Sloan JA, Loprinzi CL, Novotny PI, et al. Methodologic lessons learned from hot flash studies. J Clin Oncol 19:4280-4290, 2001.

24. Loprinzi CL, Sloan JA, Rowland KM Jr. Methodologic issues regarding cancer anorexia/cachexia trials. Research and Palliative Care: Methodologies and Outcomes, in press 2002.

25. Astin JA, Marie A, Pelletier KR, et al. A review of the incorporation of complementary and alternative medicine by mainstream physicians. Arch Intern Med 158:2303-2310, 1998.

26. Degner LF, Sloan JA. Symptom distress in newly diagnosed ambulatory cancer patients and as a predictor of survival in lung cancer. J Pain Symptom Manage 10:423-431, 1995.

27. Sloan JA, Loprinzi CL, Kuross SA, et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. J Clin Oncol 16:3662-3673, 1998.

28. Aaronson NK. Methodologic issues in assessing the quality of life of cancer patients. Cancer 67:844-850, 1991.

29. Cox DR, Fitzpatrick R. Quality of life assessment: can we keep it simple? J R Statis Soc 155:353-393, 1992.

30. Whalen GF, Ferrans CE. Quality of life as an outcome in clinical trials and cancer care: a primer for surgeons. J Surg Oncol 77:270-276, 2001.

31. Spilker B. Quality of Life and Pharmacoeconomics in Clinical Trials. New York: Lippincott Raven, 1996.

32. Lara-Munoz C, Feinstein AR. How should quality of life be measured? J Investig Med 47:17-24, 1999.

33. Leplege A, Hunt S. The problem of quality of life in medicine. JAMA 278:47-50, 1997.

34. Fayers PM, Hopwood P, Harvey A, et al. Quality of life assessment in clinical trials—guidelines and a checklist for protocol writers: the U.K. Medical Research Council experience. MRC Cancer Trials Office. Eur J Cancer 33:20-28, 1997.

35. Moinpour CM, Feigl P, Metch B, et al. Quality of life end points in cancer clinical trials: review and recommendations. J Natl Cancer Inst 81:485-495, 1989.

36. Sloan JA, Novotny P, Loprinzi CL. Analyzing quality of life endpoints in clinical trials via the SAS system, SUGL 1998, pp. 1213-1218.

37. Sloan JA, O'Fallon JR, Suman VI, et al. Incorporating quality of life measurements in oncology clinical trials. Proc Biometrics section of the American Statistical Association:282-287, 1998.

38. Sloan JA, Cella D, Frost M, et al. Assessing clinical significance in measuring oncology patient quality of life: introduction to the symposium, content overview, and definition of terms. Mayo Clin Proc 77:367-370, 2002.

39. Guyatt GR, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 77:371-383, 2002.

40. Cella D, Bullinger M, Scott C, et al. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. Mayo Clin Proc 77:384-392, 2002.

41. Sloan JA, Aaronson N, Cappelleri JC, et al. Assessing the clinical significance of single items relative to summated scores. Mayo Clin Proc 7-7:479-487, 2002.

42. Frost MH, Bonomi AE, Ferrans CE, et al. Patient, clinician, and population perspectives on determining the clinical significance of quality-of-life scores. Mayo Clin Proc 77:488-494, 2002.

43. Sprangers MA, Moinpour CM, Moynihan TJ, et al. Assessing meaningful change in quality of life over time: a users' guide for clinicians. Mayo Clin Proc 77:561- 571, 2002.

44. Symonds T, Berzon R, Marquis P, et al. The clinical significance of quality-of-life results: practical considerations for specific audiences. Mayo Clin Proc 77:572-583, 2002.

45. Juniper EF, Guyatt GH, Willan A, et al. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epid 47:81-87, 1994.

46. Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 16:139-144, 1998.

47. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 52:861-873, 1999.

48. Sloan JA, Symonds T. Health-related quality of life measurement in clinical trials: when does a statistically significant change become clinically relevant? Unpublished.

49. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Lawrence Erlbaum Associates, 1988.

50. Gould AL. Sample size re-estimation: recent developments and practical considerations. Stat Med 20:2625-2643, 2001.

51. Sloan JA, Cella D, Frost M, et al. Assessing the clinical significance of QOL measures in oncology research: state of the science—overview. Clin Ther 24:1-4, 2002.

52. Themeau TM. How many stratification factors are "too many" to use in a randomization plan? Control Clin Trials 14:98-108, 1993.

53. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31:103-115, 1975.

54. Loprinzi CL, Sloan JA, Quella SK. A phase III comparison of megesterol acetate versus dexamethasone versus fluoxymesterone for the treatment of cancer anorexia/cachexia. J Clin Oncol, 2000.

55. Shlay JC, Chaloner K, Max MB, et al. Acupuncture and amitriptyline for pain due to HIV -related peripheral neuropathy: a randomized controlled trial. Terry Beirn Community Programs for Clinical Research on AIDS. JAMA 280:1590-1595, 1998.

56. Fairclough DL. Summary measures and statistics for comparison of quality of life in a clinical trial of cancer therapy. Stat Med 16:1197-1209, 1997.

57. Fairclough DL. Missing data considerations for cancer outcomes research, COMWG Monograph, 2002.

58. Lachin JM. Statistical considerations in the intent-to-treat principle. Control Clin Trials 21:167-189, 2000.

59. Home AD, Lachenbruch PA, Goldenthal KL. Intent-to-treat analysis and preventive vaccine efficacy. Vaccine 19:319-326, 2000.

60. Stevens J. Applied Randomized Statistics for the Social Sciences (ed 3). Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

61. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of Royal Statistical Society, Methodological 57:289-300, 1995.

62. O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics 40: 1079-1087, 1984.

63. Loprinzi CL, Kugler JW, Sloan JA, et al. Venlafaxine in management of hot flashes in survivors of breast cancer: a randomised controlled trial. Lancet 356:2059-2063, 2000.

64. Loprinzi CL, Kugler JW, Sloan JA. Evaluation of coumarin for treatment of postmastectomy lymphadema. N Engl J Med 340:346-350, 1999.

65. Loehlin JC. Latent Variable Model. Hillsdale, NJ: Lawrence Erlbaum and Associates, 1992.

66. Loprinzi CL, Pisansky TM, Fonseca R, et al. Pilot evaluation of venlafaxine hydrochloride for the therapy of hot flashes in cancer survivors. J Clin Oncol 16:2377-2381, 1998.

67. Quella SK, Loprinzi CL, Barton DL, et al. Evaluation of soy phytoestrogens for the treatment of hot flashes in breast cancer survivors: A North Central Cancer Treatment Group Trial. J Clin Oncol 18:1068-1074, 2000.

68. Barton DL, Loprinzi CL, Quella SK, et al. Prospective evaluation of vitamin E for hot flashes in breast cancer survivors. J Clin Oncol 16:495-500, 1998.

69. Bloom BS, Retbi A, Dahan S, et al. Evaluation of randomized controlled trials on complementary and alternative medicine. Int J Technol Assess Health Care 16:13-21, 2000.

70. Caspi O. Integrated Medicine: Orthodox meets alternative. Bringing complementary and alternative medicine (CAM) into mainstream is not integration. BMJ 322:168, 2001.

71. Vickers A. Yes, but how do we know it's true? Knowledge claims in massage and aromatherapy. Complement Ther Nurs Midwifery 3:63-65, 1997.

72. Schrader E. Equivalence of St. John's wort extract (Ze17) and fluoxetine: a randomized, controlled study in mild-moderate depression. International Clinical Psychopharmacology 15:61-68, 2000.

73. Woelk H. Comparison of St John's wort and imipramine for treating depression: randomised controlled trial. BMJ 321:536-539, 2000.

74. Field TM, Schanberg SM, Scafidi F, et al. Tactile/kinesthetic stimulation effects on preterm neonates. Pediatrics 77:654-658, 1986.

75. Field T, Morrow C, Valdeon C, et al. Massage reduces anxiety in child and adolescent psychiatric patients. J Am Acad Child Adolesc Psychiatry 31:125-131, 1992.