ABSTRACT

Title of Document: INSTRUMENT DEVELOPMENT FOR CONTINUING MEDICAL EDUCATION EVALUATION

Jing Tian, Doctor of Philosophy, 2007

Directed By: Assistant Professor Nancy L. Atkinson
Department of Public and Community Health
University of Maryland School of Public Health

The purpose of this study was to develop a valid, reliable and adaptable CME evaluation instrument to facilitate the future CME evaluation effort as well as contribute to the literature of CME evaluation studies. A generic instrument template was first developed addressing variables in the second evaluation level based on the TPB, i.e. attitude, behavioral belief, subjective norm, perceived behavioral control and behavioral intention. The instrument was then adapted to a CME-related conference, *Preoperative Therapy in Invasive Breast Cancer: Reviewing the State of the Science and Exploring New Research Directions*. Data were collected at the conference. A total of 134 physicians returned their questionnaires. Principle axis factoring with oblique rotation was used to examine the underlying structure of the data and reduced the items in the instrument to six subscales: positive beliefs, negative beliefs, subjective norms, perceived behavioral control and behavioral intention. Factor loadings supported the existence of six valid scales. The consistency between the a priori subscales and the factors emerged served as evidence for content validity of the instrument. Overall, all the subscales had sufficient reliability (alpha>= 0.70) for early

stage instrument development showing the unidimensionality of the subscales. Scale modifications based on item analyses were conducted. The problematic items were eliminated, and the analyses were rerun. A 22-item instrument and a revised generic instrument template were finally developed. This study determined the adaptability of the theory based instrument template to the NCI CME conference and the feasibility of developing a content specific, valid and reliable CME evaluation instrument from the template assessing the changes in the concepts listed in the second evaluation level. The established and validated instrument could further be used to evaluate the effectiveness of other CME activities having the template adapted to different clinical domains addressed by each individual CME activity.

INSTRUMENT DEVELOPMENT FOR CONTINUING MEDICAL EDUCATION EVALUATION

By

Jing Tian

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Nancy L. Atkinson, Chair
Professor Barry Portnoy, Co-Chair
Professor Robert S. Gold
Professor Robert Feldman
Professor Suzanne M. Randolph

Acknowledgements

I gratefully acknowledge the contributions and advice provided by my dissertation committee members: Dr. Nancy Atkinson, Assistant Professor, Department of Public and Community Health; Dr. Barry Portnoy, Senior Advisor of Disease Prevention at the National Institutes of Health; Dr. Robert S. Gold, Dean of the School of Public Health, Dr. Robert Feldman, Professor, Department of Public and Community Health, and Dr. Susanne Randolph, Professor, Department of Family Studies.

Special thanks are given to my committee chairs, Dr. Atkinson and Dr. Portnoy. They facilitated a seamless dissertation process and provided direction, assistance, and guidance during this study. I also appreciate Dr. Atkinson for reviewing my work at every stage of this study, providing substantive recommendations, and answering my questions clearly. Particular thanks go to Dr. Portnoy who served in multiple capacities as my supervisor, mentor, and source of wisdom and calm. I am grateful to him for believing in my research ideas and suggesting that I conduct this study.

I wish to express thanks to Dr. Barnette Kramer, Associate Director for Disease Prevention at the National Institutes of Health, who provided me with expert and timely feedback on the instrument and the template from a clinical perspective. In addition, Dr. Kramer and Dr. Portnoy helped me to obtain National Institutes of Health funding for this dissertation, which not only secured the research financially, but made it more valuable and meaningful for me.

I give special thanks to Ms. Sylvia K. Scherr, former Executive Director of Continuing Medical Education Office, who connected me with Dr. Jo Anne Zujewski, Senior Medical Oncologist of NCI, the meeting organizer of the pre-operative breast cancer therapy

conference where I collected my data. Heartfelt thanks are due to Dr. Zujewski whose commitment to improving continuing medical education and genuine interest in understanding the physicians' attitudes and confidence change after the conference allowed this study to be conceived. Dr. Zujewski also provided me with enormous help in reviewing the template and draft instruments, referring participants for cognitive testing of the instrument, and facilitating data collection. Without Dr. Zujewski's support and invaluable suggestions, this study would not have been possible.

I wish to thank Dr. Randolph, for suggesting the inclusion of a cognitive testing procedure. This proved to be a critical step for the development of the (insert name of instrument) instrument. I wish to express my gratitude to Dr. Gold, for serving as the psychometrician on my committee, providing guidance for my data analysis and results interpretation. I would like to acknowledge Dr. Feldman for serving as the behavioral theorist on my committee, advising the rational of the underlying theory and interpreting the unexpected constructs revealed by my data analysis. My thanks extend to the over 170 clinicians who volunteered as participants in this study and the four meeting planners at Palladian Partners, Inc. who assisted me with data collection and gave me moral support in my effort to assertively collect data. Particular thanks to my peer Kantahyanee Murray, who helped me practice my presentation and provided helpful feedback. This support psychologically prepared me for my dissertation defense meeting.

Particular thanks go to my mother Ms. Suorong Li and my father Mr. Zhi Tian for staying with me in the U.S. for six months, providing encouragement, and enthusiasm throughout the dissertation process, and preparing the most wonderful meals in the world everyday. The emotional, physical, and culinary support provided by my parents made it

easier to accomplish completing my dissertation amidst the challenges of living in a foreign country.

It is with immense love and adoration that I thank my husband, Xianghua Yin, whose constant care, advice, concern, devotion, and of course, fantastic cooking, enabled me to complete this dissertation project in record time.

LIST OF TABLES

TABLE OF CONTENTS

Chapter 1: Introduction

Statement of the Problem

In a recent systematic review by Tian and colleagues, the authors concluded the need for the development of the valid and reliable CME evaluation instrument.

> "A standard questionnaire with core items on attitudes, self-efficacy, and beliefs that can be adapted for different CME programs for evaluation and comparison is needed to enable the comparison of effectiveness across different CME interventions. Comparison of these standardized results will help researchers understand factors influencing the effectiveness of different CME programs and guide future intervention design. The concepts being evaluated in this standard questionnaire should include but not be limited to attitudes, beliefs and self-efficacy. The trunk of the items assessing those concepts would be the same with content area being specified according to different clinical domains." (Tian et al, 2007)

As a result of the above cited review, the purpose of this study was to develop a valid, reliable and adaptable CME evaluation instrument to facilitate future CME evaluation efforts as well as contribute to the literature on CME evaluation studies.

As the official accrediting body for CME programs, Accreditation Council for Continuing Medical Education (ACCME) identifies, develops, and promotes the standards for quality CME that physicians could use for the maintenance of their competence and the incorporation of new knowledge for the purpose of improving quality medical care for patients and their communities (ACCME, 2006, p. 1). According to ACCME, CME includes educational activities that aim at maintaining, developing, or increasing the knowledge, skills, and professional performance and relationships used by a physician to

provide services for patients, the public, or the profession. The body of knowledge and skills generally recognized and accepted by the profession as within the basic medical sciences, the discipline of clinical medicine, and the provision of health care to the public is the content of CME (ACCME, 2006, p. 1).

The ultimate goal of CME programs is to enhance the quality of patient care available in the United State as well as other regional areas through professional education. Physicians spend a considerable amount of time in CME to maintain their medical licenses. According to State Medical Licensure Requirements and Statistics (2006), forty-seven of fifty-four state and territorial medical licensing boards require completion of 12 to 50 hours of CME per year.

CME activities are underpinned by a belief that knowledge gains lead physicians to improve their medical practices and patient outcomes (Davis, 1999). Many reviews have been published during the last decade trying to summarize CME evaluation studies to assess their effectiveness (Davis et al., 1999; Hogan et al., 2001; Cauffman et al., 2002; Thomson O'Brien et al., 2001; Bloom et al., 2005; Amin, 2000; Jaussent et al., 2004; Wutoh et al., 2004; Curran et al., 2005, Davis et al., 2006). Previous reviews of CME evaluations have shown that the questionnaires used in the CME evaluation studies have generally lacked a theoretical background which may have resulted in misleading interpretations of study results; this also limited comparisons across different CME evaluations (Jaussent et al., 2004).

According to a recent released evidenced report on the Effectiveness of Continuing Medical Education by the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (USHHS), the reliability and validity of the

instruments that have been used to assess CME effectiveness limited the evidence. Consistent with this result, in a recent review of 32 randomized clinical trials of CME evaluation studies, Tian and colleagues (2007) demonstrated varied questionnaires, surveys, and scales were used to evaluate outcomes of CME, including: physicians' knowledge, beliefs, attitudes, perceived confidence; patients' satisfaction of consultations, perception of communication skills; and depression level. Half of the 32 studies (N=16) used questionnaires specific to the clinical domains addressed. Six of these 16 studies (18.8%) adapted existing instruments and provided reliability and validity information. Ten of the remaining studies (31.3%) developed their own instruments. However, none of the studies using self-developed instruments documented reliability or validity information.

After assessing participant satisfaction with CME, both attitude and knowledge change need to be evaluated to assess whether the determinants for physician behavior change are in place (Kirkpatrick, 1994). These measures can serve as proxy measures for physician behavior change until more rigorous evaluation methods can be implemented. Tian et al.'s study (2007) found a lack of valid and reliable CME instruments; this indicated that a standard questionnaire with core items on attitudes, self-efficacy, and beliefs that can be adapted for different CME programs for evaluation and comparison is needed to enable the comparison of effectiveness across different CME interventions. This information would allow the results to be adequately interpreted and compared, and other researchers would be able to assess the adequacy of the measurements used.

Comparison of these standardized results will help researchers understand factors influencing the effectiveness of different CME programs and guide future intervention

design. The concepts being evaluated in this standard questionnaire should include but not be limited to attitudes, beliefs, and self-efficacy. The trunk of the items assessing those concepts would be the same with content area being specified according to different clinical domains.

Donald Kirkpatrick developed a four-level outcome evaluation model in 1994. It has been widely used in assessing training effectiveness. According to this model, highly effective training programs should result in four kinds of outcomes. The four levels of outcome evaluation in the model are level 1 evaluation—reaction; level 2 evaluation—learning; level 3 evaluation—behavior, and level 4 evaluation—results. Curran and Fleet adapted Kirkpatrick's model for the field of CME in 2005. Explained in the context of CME, the four levels of outcome evaluation in this model are learner satisfaction (level 1); learning outcomes (level 2); performance improvement (level 3); and improved patient/health outcomes (level 4). In addition, according to this model, evaluation should always begin with level one, and then, as time and budget allows, should move sequentially through levels two, three, and four. Information from each prior level serves as a basis for the next level's evaluation. Thus, each successive level represents a more precise measure of the effectiveness of the training program but, at the same time, requires a more rigorous and time-consuming analysis.

The Theory of Planned Behavior (TPB) has long been used to investigate physicians' changing their clinical practices (Tian et al., 2007) According to TPB, in the context of CME, physicians' intention to perform certain clinical practices is determined by their attitude toward performing these clinical practices, associated subjective norms, and perceived behavioral control of performing these clinical practices. TPB provides a

detailed structure among the concepts that the proposed instrument is trying to measure. In addition, the constructs from the Social Cognitive Theory (SCT), e.g. outcome expectation and self-efficacy, are more often measured in CME evaluation studies. Therefore, the constructs of the TPB that are parallel to the ones in the SCT were used as proxy measures in this research, i.e. perceived behavioral control in TPB served as a proxy measure of self-efficacy in SCT; behavioral belief in TPB served as a proxy measure of outcome expectation in SCT.

Constructs of attitudes, subjective norms, and perceived behavioral control from the TPB fit well into the second evaluation level of the Kirkpatrick model (1994). Behavioral intention is determined by the constructs from the second level while predicting the element (behavior) in the third evaluation level. Therefore, it is located between the second and third evaluation levels. Behavioral intention questions would serve as a proxy measure of physician behavior in the third evaluation level.

A variety of evaluation measures were used in previous CME evaluations and included qualitative and quantitative strategies (Tian et al, 2007). However, among all the quantitative questionnaires being used, few documented validity and reliability information (Tian et al, 2007); even fewer addressed the variables in the second evaluation level (Tian et al, 2007). This limited comparisons of effectiveness across CME courses.

Purpose of the Study

The purpose of conducting this study was to create a theoretically driven, valid, reliable, and adaptable CME evaluation instrument addressing attitudinal determinants of physician behavior change, i.e. attitudes, beliefs, subjective norms, perceived behavioral control (self-efficacy), and behavioral intention. Goals of the study were to: (1) describe the development of a CME evaluation tool that could serve as a model across CME courses; (2) apply the Theory of Planned Behavior (TPB), Social Cognitive Theory (SCT), and Kirkpatrick's evaluation model to CME evaluation; (3) evaluate reliability and validity of the proposed CME instrument; (4) describe how this instrument could be used to facilitate future CME evaluation efforts; and (5) enable the evaluation of physician behavior change (intention) as the result of CME.

In summary, this dissertation described the background, theoretical bases, and the specific steps for developing the instrument and examining its validity and reliability. The content area for this model instrument was preoperative therapy for breast cancer. The constructs this instrument was trying to measure were (1) behavioral beliefs (outcome expectations) of performing a certain clinical practice; (2) attitudes toward performing a certain clinical practice; (3) subjective norms of performing a certain clinical practice; (4) perceived behavioral control (self-efficacy) in performing a specific clinical practice; and (5) behavioral intention to perform a specific clinical practice. The primary focus of this study was to evaluate the validity and reliability of this CME instrument.

Research Questions

TPB suggests that measuring behavioral intention would require at least the following three subscales necessary to measure each element of the predictive model with a single subscale:. attitudes, subjective norms, and perceived behavior control. Given the need to develop such an instrument to more effectively evaluate CME outcomes, the following questions guided this instrument development study:

1. Will a thorough content validation process satisfy the needs for instrument validity?
2. Will a psychometric examination of the draft instrument reveal any unexpected measurement subscales?
3. Can an instrument for this purpose be developed with subscales consistent with the theoretical domains?
4. Can an instrument be developed with acceptable levels of reliability for each of the necessary subscales?
5. Will a thorough instrument development process result in an instrument that is appropriate for evaluation of CME?

Specific Aims

In order to the answer the research questions of the study, the following specific aims were developed to guide the research and data analysis. The specific aims of the study were to:

1. Construct conceptual definitions for each of the theoretical concepts.
2. Identify the best measurement and item strategies for each of the needed items.
3. Determine the face validity by having experts review the initial item pool.

4. Conduct cognitive testing of the instrument with the target audience to determine item understandability and acceptability.

5. Pilot test the instrument with the target audience to track the length of time spent, item understandability, and acceptability.

6. Provide the evidence of content validity for the proposed instrument.

7. Collect pretest survey data from participating physicians in the Preoperative Therapy in Invasive Breast Cancer conference.

8. Estimate the sample size sufficiency of the study.

9. Determine the subscales/underlying factors for the proposed instrument by conducting exploratory factor analysis with oblimin rotation.

10. Compare the conceptual definitions of the subscales/underlying factors with the underlying theoretic domains of the proposed instrument for unexpected measurement subscales.

11. Investigate the percentage of variance of the proposed instrument has been explained by the factor(s) extracted.

12. Examine the construct validity for the proposed instrument.

13. Evaluate the internal consistency reliability of each subscale/factor for the proposed instrument.

14. Finalize the psychometrically acceptable draft of an instrument that is consistent with current theory.

Rationale for the Study

The reliability and validity of the instruments that have been used to assess CME effectiveness limited the evidence (AHRQ, 2007). A valid, reliable, and standard questionnaire with core items on attitudes, beliefs, and self-efficacy that can be adapted for different CME programs for evaluation and comparison was needed to enable the comparison of effectiveness across different CME interventions. Comparisons of these standardized results can help researchers understand factors influencing the effectiveness of different CME programs and guide future intervention design. It can also facilitate future NIH CME evaluation efforts (Tian et al, 2007).

Summary

Chapter 1 presented the background of the study, the purpose of developing an instrument to measure the variables in the second level of Kirkpatrick's evaluation model, research questions, definitions for key terms, the significance and rational for the study and delimitations. Chapter 2 provides a review of relevant literature and contextual factors justifying this study, framework for assessing the quality of research methods, theoretical basis for the developing the proposed instrument, and the existing attitude scales for CME evaluation studies. Chapter 3 provides the methods used and processes for instrument development and refinement, data collection, and data analysis. Chapter 4 provides the results of cognitive testing, pilot testing, construct validity, reliability and item analyses. Chapter 5 provides the discussion of the study results, limitations of the research, recommendations for future research, and conclusions.

Definition of Terms

Operational definitions of many terms used throughout this study were addressed as follows:

Continuing Medical Education: Educational activities which serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a physician uses to provide services for patients, the public, or the profession (ACCME, 2006, p. 1).

Impact Evaluation: Assessment of the immediate effects of a program, such as knowledge change after participating in the health education program (McDermott et al., 1999).

Outcome Evaluation: Assessment of the long-term effects of a program, in terms of the morbidity and mortality rates. The objective of outcome evaluation is to validate the results being achieved and the reasons and strategies being used for their achievement (Evaluation Office United Nations Development Program, 2002).

Attitudes: The degree to which performance of the behavior is positively or negatively valued. Attitudes toward a behavior is determined by the total set of accessible behavioral beliefs linking the behavior to various outcomes and other attributes. Specifically, the strength of each belief is weighted by the evaluation of the outcome or attribute, and the products are aggregated (Fishbein & Ajzen, 1975).

Behavioral Beliefs: The subjective probability that the behavior will produce a given outcome. It is assumed that these accessible beliefs—in combination with the subjective values of the expected outcomes—determine the prevailing attitude toward the behavior. Specifically, the evaluation of each outcome contributes to the attitude in direct proportion to the person's subjective probability that the behavior produces the

outcome in question (Fishbein & Ajzen, 1975). This could serve as a proxy measure for outcome expectation by Bandura (1977b, 1986).

Behavioral Intention: An indication of a person's readiness to perform a given behavior, and it is considered to be the immediate antecedent of behavior. The intention is based on attitudes toward the behavior, subjective norms, and perceived behavioral control, with each predictor weighted for its importance in relation to the behavior and population of interest (Fishbein & Ajzen, 1975).

Perceived Behavioral Control: People's perceptions of their ability to perform a given behavior (Fishbein & Ajzen, 1975). This could serve as a proxy measure of self-efficacy by Bandura (1977b, 1986)

Subjective Norms: a person's subjective norms is determined by his normative beliefs—whether important referent individuals approve or disapprove of performing the behavior, weighted by his motivation to comply with those referents (Fishbein & Ajzen, 1975).

Outcome Expectation: Anticipatory aspects of behavior, i.e. antecedent determinants of behavior (Bandura, 1977b, 1986).

Outcome Expectancy: The values that a person places on a particular outcome. (Bandura, 1977b, 1986).

Self-Efficacy: The confidence that participating physicians have for performing the preoperative therapy in invasive breast cancer (Bandura, 1977b, 1986).

Physician: A medical doctor who graduated from an U.S. accredited medical school or a board certified foreign trained medical doctor.

Chapter 2: Review of the Literature

Introduction

This chapter presents a review of the literature on continuing medical education (CME), related evaluation issues, and the theory base for CME evaluation. This chapter first defines and describes what continuing medical education is and demonstrates its purpose and the requirements of medical licensure for physicians nation-wide. The importance of conducting evaluations and different categories of evaluations are discussed next. Thirdly, the review discusses the underlying theoretical basis for the proposed instrument development study—Kirkpatrick's Model for outcome evaluation and its application in the field of continuing medical education. The section after is a review of the previous continuing medical education evaluation research providing the state-of-the-art information on CME evaluations. Several conclusions drawn from this review were consistent with other recent review studies. Next, the Theory of Planned Behavior (TPB) and Social Cognitive Theory (SCT) are thoroughly reviewed. These two health behavior theories serve as the underlying theoretical frameworks for constructs being included in the proposed continuing medical education evaluation instrument.

This review also examines existing attitude scales for CME evaluation to inform the content and process of the proposed scale development study. Since only three CME instruments from our previous review study addressing the variables in the second evaluation level (attitudes, knowledge, self-efficacy or beliefs) have documented information for reliability and validity, each of them is discussed individually. Another instrument (Diabetes Attitude Scale) is also discussed here, as it is a well-validated attitude scale with demonstrated reliability. Another instrument (Primary Care

Practitioner's Attitudes and Confidence Scale) is included to provide a review of promising strategies used in scale development.

Continuing Medical Education

Definition, Purpose, and Requirements for Medical Licensure

The Accreditation Council for Continuing Medical Education (ACCME) is the official accrediting body for Continuing Medical Education programs (ACCME, 2006, p. 1). It includes representatives from the American Board of Medical Specialties, the American Hospital Association, the Association of American Medical Colleges, the Association for Hospital Medical Education, the Council of Medial Specialty Societies, and the Federation of State Medical Boards.

The mission of ACCME is to identify, develop, and promote the standards for quality continuing medical education (CME). Physicians utilize these standards to maintain their competence and to incorporate new knowledge in order to improve quality medical care for patients and their communities (ACCME, 2006, p. 1). This mission has been fulfilled through a voluntary system, which regulates accrediting CME providers and a peer-review process responsive to changes in medical education and the health care delivery system.

The primary responsibilities of the ACCME include the setting and administration of the standards and criteria for quality CME providers for physicians and related professionals; certification of the accredited providers' capability to meet the requirements of the essential areas; relation of CME to the continuum of medical education and medical care; evaluation of the effectiveness of CME policies; assistance to CME providers to improve their programs continually; and assurance of physicians, the public, and the CME

community with the fact that CME programs meet the ACCME's criteria for compliance with the essential areas (ACCME, 2006, p. 1). According to ACCME, Continuing Medical Education (CME) "consists of educational activities which serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a physician uses to provide services for patients, the public, or the profession. The content of CME is that body of knowledge and skills generally recognized and accepted by the profession as within the basic medical sciences, the discipline of clinical medicine, and the provision of health care to the public" (ACCME, 2006, p. 1).

The broad definition of CME above implies that CME includes all continuing educational activities that assist physicians in carrying out their professional responsibilities more effectively and efficiently. Based on this definition, physicians responsible for managing a health care facility could consider a course in management as appropriate CME. Likewise, a course in educational methodology and a course in practice management would be appropriate CME for physicians teaching in a medical school and practitioners interested in providing better service to patients, respectively. Some continuing educational activities in which physicians engage are excluded from CME. These include continuing educational activities that are not related directly to physicians' professional work or continuing educational activities that respond to a physician's non-professional educational need or interest--such as personal financial planning.

Offices in different government health agencies, hospitals, and medical institutes provide different CME programs in order to meet different needs of their physicians. Each CME program has a specific purpose according to the different agencies providing them

and the different medical/health care domains they address. The ultimate goal of CME programs is to enhance the quality of patient care available in the United State as well as other regional areas through professional education. According to the Preamble to the 2004 Updated ACCME Standards for Commercial Support (ACCME, 2004), as well as CME activity guides of different medical institutes, the purposes of CME in general are as follows:

1. Providing information and opportunities for physicians and other health care providers to learn about more effective, up to date, and efficient health care delivery strategies in new health care markets.

2. Providing information and opportunities to develop knowledge bases and skills according to the latest technological and scientific advancements in medicine.

3. Providing continuing medical educational programs that meet the criteria as designated by the American Medical Association's (AMA) Physician's Recognition Award (PRA) and following the essentials, guidelines, and standards of the ACCME.

According to Continuing Medical Education for Licensure Reregistration (2006, p. 47), fifty-eight boards (including Wyoming, as of January 2007) require from 12 hours (Alabama) to 50 hours (several states) of continuing medical education (CME) per year for license reregistration. CME content—such as HIV/AIDS, risk management, or end of life palliative care—has been mandated by some states. In addition, many states require that a specific percentage of CME should be Category 1 Credit™.

Educational activities must be planned by an accredited provider, and the activities must meet following criteria in order to be designated for AMA PRA Category 1 Credit. They must be consistent with AMA definition of CME; content provided must be

appropriate for a physician audience; they should be consistent with both the relevant Council on Ethical and Judicial Affairs (CEJA) opinions, as well as the ACCME standards for commercial support, and are non promotional in nature; they should address demonstrated educational needs and communicate a clearly identified educational purpose and/or objectives; they should use learning methodologies and format(s) appropriate to the activity's educational purpose and/or objectives; they should use evaluation mechanisms to assess an activity's quality and relevance to its purpose and/or objectives; they should include a means for the provider to record the actual credits claimed by each physician participant; they are designated for AMA PRA Category 1 Credit in advance but not afterwards; they should include the designation statement in any activity materials that reference CME, with the exception of "save the date" or similar notices (AMA Physician's Recognition Award and credit system, 2006, p. 4-5).

Since 1968, CME participating physicians have been recognized by the AMA with the PRA. The AMA PRA Category 1 Credit™ system has become the CME standard for licensing boards and specialty organizations nationwide. Physicians should participate in at least 50 credits per year of educational activities that meet AMA standards in order to earn the PRA (ACCME, 2006, p. 1). Although AMA PRA Category 1 Credit™ is recognized by all U.S. jurisdictions, the PRA or the PRA application is accepted by 44 jurisdictions as proof of having met those requirements. PRA is accepted by 34 medical boards as proof of completion of their CME requirement now, which simplifies the medical relicensure process.

Importance of Evaluation

According to McDermott and colleagues (1999), evaluation is assessing the utility of implementation plan and procedures, the extent and quality of implementation, and the effects of implementation on immediate learning outcomes. The term "program" may include any organized action such as media campaigns, service provision, educational services, public policies, research projects, etc. (Center for Disease Control and Prevention [CDC], 1999).

Evaluations should produce credible, relevant, timely, and objective findings and conclusions on program performance. The findings should be based on valid and reliable data collections and analysis. Evaluations should present these findings and conclusions in a clear and balanced manner by indicating the reliability of the findings (Treasury Board of Canada, Secretariat, 1998).

Various types of evaluation are used to assess different aspects or stages of program development. Terminology and definitions of evaluation types are not uniform; a number of types being used widely are as follows:

Formative Evaluation is defined as the ongoing process of evaluation while the program is being developed and implemented. It is also called process evaluation. Its primary goal is improving the program. Two important elements of the formative evaluation are quality assurance and control (McDermott et al., 1999).

Process Evaluation is defined as examination of the intervention itself and the degree to which it was implemented as planned and necessary (Thomas, et al. 2000).

Summative Evaluation is the assessment of the degree to which the prespecified objectives have been made by the project or the degree to which the project was useful for

the target population (McDermott et al., 1999). Quantitative approaches are often used by summative evaluations. It is often conducted by an outside evaluator for the purpose of objectivity. Impact evaluation and outcome evaluation are two forms of summative evaluation.

Impact Evaluation is defined as the assessment of the immediate effects of a program, such as knowledge change after participating in the health education program (McDermott et al., 1999).

Outcome Evaluation is defined as assessment of the long-term effects of a program, in terms of the morbidity and mortality rates. According to the Evaluation Office United Nations Development Program (2002), the objective of outcome evaluation is to validate the results being achieved and the reasons and strategies being used for their achievement. Outcome evaluations investigate how and why outputs and strategies contributed to achievement of outcome by comparing the planned with intended outcome achievement. The focused questions of the evaluation are relevance, effectiveness, sustainability, and impact. The outcome achievements are evaluated by comparing indicators before and after the intervention. It relies on monitoring data on information form external sources. The outcome evaluation always conducted in a time-bound, periodic and in-depth manner by external evaluators and partners if possible. The evaluation findings provide program managers with strategy and policy options; it also provides a basis for learning and demonstrates accountability.

Kirkpatrick's Model for Outcome Evaluation

The four-level outcome evaluation model developed by Donald Kirkpatrick (1994) has been widely used in assessing training effectiveness. According to Kirkpatrick, highly effective training program should result in four kinds of outcomes. The four levels of outcome evaluation in the model are level 1 evaluation—reaction; level 2 evaluation—learning; level 3 evaluation—behavior, and level 4 evaluation—results.

The goal of level 1 evaluation is to measure participants' immediate reactions to the training program. In addition to the measurement of overall customer satisfaction, it should also include measurement of participants' reactions or attitudes toward specific components of the program, such as the instructor, the topics, the presentation style, the schedule, and audiovisuals. Each of these components is composed of several sub-components for evaluation. In short, level 1 evaluation is much more than just the satisfaction with the training program.

It is important to evaluate participants' reactions to the training program. Level 2 outcomes (learning) and level three outcomes (transfer of learning) are only able to occur when participants have positive attitudes toward the training program. Therefore, positive reactions are important as unpopular training programs are likely to go unattended and to be eliminated. Finally, the measurement of specific aspects of the training program can provide important information about the aspects for future improvement of the training program.

The goal of level 2 evaluation is to determine what the participants learned immediately during the training program. One should expect to find clear learning outcomes according to the specific learning objectives of the instructors. Learning

outcomes can include changes in knowledge, skills, or attitudes. Knowledge is usually measured by the achievement tests constructed by the instructor (i.e., tests designed to measure the degree of learning that has taken place. A performance requiring the test taker to create a product or demonstrate a process is used to evaluate the skills obtained. Attitudes are usually measured with questionnaires. The advantages of level 2 evaluations include 1) helping trainers improve their training program by demonstrating participants' learning outcomes; and 2) providing a basis for interpreting the results of level 3 evaluations.

The goal of level 3 evaluation is to find out the on-the-job-behavior (OJB) changes of the participants due to their participation in the training program. It specifically involves measuring the transfer of knowledge, skills, and attitudes from the training environment to the workplace. However, level 1 and level 2 outcomes are equally important because participants need to react positively to the training program (level 1), and they need to learn the material (level 2) in order to be motivated and able to apply what they have learned in their workplace.

Behavior changes at the workplace are often harder to measure than reaction and learning immediately after the training, so level three is always harder than level one and level two evaluations to assess (Kirkpatrick, 1994). In addition, one should also allow time for behavior transfer and data collection at the workplace (Kirkpatrick, 1994). The advantages of level three evaluations include 1) measuring actual on-the-job behavior changes; 2) determining outcomes that are the intervening variables or factors necessary for the outcomes in the fourth level; and 3) when combined with positive results in levels 1 and 2, providing sufficient evidence of the merit and usefulness of a training program.

The goal of level 4 evaluation is to find out if the training program led to final results. In the Kirkpatrick model, level four outcomes are either changes in financial outcomes (such as positive return on training investment or increased profits) or changes in variables that should have a relatively direct effect on financial outcomes at some point in the future.

Several challenges make it difficult to establish firm evidence that a training program was the key or only source that produced level four outcomes (Kirkpatrick, 1994). First, it takes time for those outcomes to occur after the training program. Other factors may also occur during that time period that confound the reason. Second, the additional causal variables operating on the same level make the isolation of the training effect very difficult. Third, level four outcomes are usually more distal instead of proximal training outcomes. Thus, the evidence obtained from level four evaluation is usually not as strong as those from lower level evaluations, especially levels 1 and 2 which are relatively easy to document.

In summary, evaluation should always begin with level 1, and then, as time and budget allows, should move sequentially through levels 2, 3, and 4. Information from each prior level serves as a basis for the next level's evaluation. Thus, each successive level represents a more precise measure of the effectiveness of the training program but, at the same time, requires a more rigorous and time-consuming analysis.

Modified Kirkpatrick's Model in CME Outcome Evaluation

Curran and Fleet (2005) adapted Kirkpatrick's model for the field of CME. Outcomes in each evaluation level are explained in the context of CME. The four levels

of outcome evaluation in the model are learner satisfaction (reaction, level 1); learning outcomes (learning, level 2); performance improvement (behavior, level 3); and patient/health outcomes (results, level 4).

Level 1 evaluation is intended to evaluate how well participants liked a CME program. It generally provides data concerning participants' perceptions of and satisfaction with program objectives, content, instruction, delivery, and instructors. Level 2 evaluation involves some form of assessment of changes in skills, knowledge, or attitudes among learners; it is most commonly conducted through pre- and post-test study designs. Level 3 evaluation provides information on the extent to which learning has influenced the post-learning behavior or performance of a learner in their practice setting. Evaluating at this level attempts to answer the question: are the newly acquired skills, knowledge, or attitudes being used in the everyday environment of the learner? Level 4 of evaluation is concerned with measuring tangible results that are influenced by the performance of the learner as a result of participation in the continuing education activity. These tangible results can be transferred to a health perspective (e.g. improving patient health or improving efficiencies). Evaluation at this level is challenging given the variety of uncontrollable variables a learner encounters when he or she leaves the boundaries of a program.

Previous Research on CME Evaluation

Many reviews have been published during the last decade trying to summarize CME evaluation studies. Several key findings were demonstrated in those reviews, which were also quite consistent with each other.

According to Cantillon et al. (1999), continuing medical education for general practitioners should be largely based on the work they do. The targeted behaviors, baseline compliance, the characteristics of the CME interventions, and the results compose a great complexity and substantially vary (Thomson O'Brien et al, 2001). The heterogeneity of CME evaluation results has been explained best by differences in the interventions (Thomson O'Brien et al, 2001).

Amin et al. (2000) argued that physician learning is such a distinct phenomenon that it is highly inclined towards autonomy and self-directed learning. The more successful CME interventions are those that are modeled upon a solid theoretical background, tailored towards individual learning needs and preferences, and concentrated on educational learning components. Many widely practiced CME interventions that did not follow the above principles turned out to be ineffective. Many reviews (Davis, et al., 1999; Hogan et al., 2001; Cauffman et al., 2002; Thomson O'Brien et al., 2001; Bloom et al., 2005) have found that: 1) lectures and unsolicited printed material are weak forms of CME; 2) didactic sessions alone are unlikely to change professional practice; 3) small-group interactive CME and problem-based interventions that provides practice opportunities appears to be the most effective strategies in changing physician behavior and can result in relatively large changes in professional practice; and 4) significant event audits, peer review, group based learning, and reminders by computer have all been shown to be effective educational strategies for general practice.

Recently, the Internet has grown in popularity as a medium for knowledge transfer, and it has become an important CME channel (Curran et al., 2005). According to Sklar et al. (2001), as Internet usage has grown, the number of web-based CME providers and the

number of CME websites has increased significantly. In 2000, 96 CME sites were available; this number increased to 200 by 2001 (Sklar et al., 2001). Therefore, there is a need to examine in further detail the nature and characteristics of those web-based learning technologies, environments, and systems across different clinical disciplines that are most effective in changing physicians' practices and, ultimately, in improving patient and health outcomes.

The majority of evaluation research on web-based CME is still based on the participant satisfaction level. Wutoh et al.'s study (2004) demonstrated that Internet-based CME programs were just as effective as traditional formats of CME interventions in physicians' knowledge change. However, little is known about the transferability of those positive knowledge changes into clinical practice changes. Limited research has demonstrated performance change in clinical practices, and no studies reported in the literature have yet demonstrated the effectiveness of web-based CME in impacting patient or health outcomes (Curran et al, 2005).

The reviews of the CME evaluation literature also revealed considerations when assessing physician performance. One found that both qualitative and quantitative strategies could be used in assessing the impact of CME interventions on physician clinical performance (Jaussent et al., 2004). Qualitative methods include observation, focus groups, and individual interviews. However, these methods require specific skills, are time-consuming, and are difficult to implement on a large scale (Jaussent et al., 2004). In contrast, quantitative methods use survey questionnaires or standardized instruments and are more appropriate for large scale or repeated evaluations.

Evaluating knowledge, beliefs, attitudes, and self-efficacy (variables in the second evaluation level) is relatively proximal to the intervention and easier to document (Kirkpatrick, 1994). Despite the relative ease in conducting this level of evaluation, a previous review found that many methodological factors could affect the accuracy of ratings provided by these instruments, including the quality of the instrument in terms of validity, reliability, and sensitivity to changes (Jaussent et al., 2004). This review found that properties of validity, reliability, and sensitivity of questionnaires designed to assess the knowledge, perceptions, and practices of health care professionals with regards to alcoholic patients were often neglected. This situation was similar in other clinical and health care domains. Moreover, these questionnaires generally lacked a theoretical background. Hence, the interpretation of responses to those questionnaires may have been misleading. In order to assure that minimum evidence of validity, reliability, and sensitivity are available, this review recommended that journals require that authors provide this information when reporting survey results, questionnaire development, or standardized instruments. A standardized survey questionnaire following a strict methodology should be developed for teams that train medical staff caring for patients.

Jaussent et al.'s conclusion (2004) was consistent with a recent review by Tian et al. (2007) and could be extended to the entire scope of CME addressing different clinical domains. In CME evaluation studies published 2000 through 2006, the questionnaires designed to assess the knowledge, perceptions, attitudes, beliefs as well as self-efficacy did not demonstrate a solid theoretical basis, and their validity and reliability information was often not documented. These findings pointed to the need for the development of a validated and reliable standardized survey questionnaire following a strict methodology.

Such an instrument could be developed to contain core items addressing the aforementioned concepts that could be adaptable to different clinical domains, enabling the comparison of the outcomes among different CME interventions.

When evaluating clinical practices (variables in the third evaluation level), physicians have limited ability to accurately self-assess and often tend to over-estimate the positive changes they make. Learning needs assessment and professional competence evaluation should therefore focus more on objective, external assessment than self-reported questionnaires. Three other systematic reviews of CME evaluation studies (Cantillon et al., 1999; Wutoh et al., 2004; Davis et al., 2006) drew this same conclusion.

According to previous reviews (Faber et al, 2005; van Zyl et al, 2004; Gask et al, 2004; Razavi et al, 2003; Bland et al, 2003; Flores et al, 2002; Ray et al, 2001; Gielen et al, 2001; Curtis et al, 2000; Martling et al, 2000; Haug et al, 2000; Thompson et al, 2000), evaluating patients or health outcomes (variables in the fourth level) is very difficult in terms of attributing the results solely to the training program. Other factors and causal variables operating on the same level may also occur during the time period for health outcomes to develop after the CME intervention. Those variables will confound the outcomes and make it hard to isolate training effects. In addition, training outcomes in this level are usually more distal than proximal, so that the evidence obtained from this level of evaluation is usually hard to document and is not as strong as those from lower level evaluations.

Table 2-1 CME Evaluation Review Studies

Author, Year Published	Review Focus	Time Frame of the Articles	Selection Criteria/Delimitations
Davis DA, Mazmanian PE, Fordis M et al., 2006	Accuracy of physician self-assessment compared with observed measures of competence: a systematic review	MEDLINE (1966-July 2006), EMBASE (1980-July 2006), CINAHL (1982-July 2006), PsycINFO (1967-July 2006), the Research and Development Resource Base in CME (1978-July 2006), and proprietary search engines were searched using terms related to self-directed learning, self-assessment, and self-reflection 17 met all inclusion criteria	Studies were included if they compared physicians' self-rated assessments with external observations, used quantifiable and replicable measures, included a study population of at least 50% practicing physicians, residents, or similar health professionals, and were conducted in the United Kingdom, Canada, United States, Australia, or New Zealand. Studies were excluded if they were comparisons of self-reports, studies of medical students, assessed physician beliefs about patient status, described the development of self-assessment measures, or were self-assessment programs of specialty societies. Studies conducted in the context of an educational or quality improvement intervention were included only if comparative data were obtained before the intervention.
Bloom BS, 2005	A review of systematic reviews	1/1/1984-10/30/2004 26 reviews met inclusion criteria	English-language, peer-reviewed meta-analyses and other systematic reviews of CME programs that alter physician behavior and/or patient outcomes.
Curran VR, Fleet L, 2005	A review of evaluation outcomes of web-based continuing medical education	A search of Medline using the Mesh terms “Internet” and “continuing medical education” was conducted for all years up to December 2003. Reference sections of the studies were also reviewed for additional peer reviewed literature.	Studies were included in the review if they included at least one level of evaluation as described by Kirkpatrick, involved the use of the Internet as a medium for delivering a structured program or course of CME study, and included a doctor audience.

Wutoh R, Boren SA, Balas EA., 2004	eLearning: a review of Internet-based continuing medical education	MEDLINE (1966 -1/2004), CINAHL (1982 -12/2003), ACP Journal Club (1991-8/2003), and the Cochrane Database of Systematic Reviews (third quarter, 2003). 16 studies met the eligibility criteria	Studies were included in the analyses if they were RCTs of Internet-based education in which participants were practicing health care professionals or health professionals in training. CME interventions were categorized according to the nature of the intervention, sample size, and other information about educational content and format.
Jaussent S, Labarere J, Boyer JP et al., 2004	Psychometric characteristics of questionnaires designed to assess the knowledge, perceptions and practices of health care professionals with regards to alcoholic patients	1/1/1964 to 12/31/2002 A total of 57 relevant publications involving 39 original instruments were identified	The aim of this paper is to describe the properties of French and English language questionnaires designed to assess the knowledge, perceptions, and practices of health care professionals with regards to alcoholic patients.
Cauffman JG, Forsyth RA, Clark VA et al., 2002	Randomized controlled trials of continuing medical education: what makes them most effective?	Published in the10-year period from 1982 to 1991. 20 randomized controlled trials (RCT) were identified.	Selected CME studies that met predetermined criteria, as described in the earlier work by Davis et al., were examined in this study. They had to have family physicians and/or general practitioners as part or all of the physician sample. Examine the effect physician performance and/or patient health care outcomes.

Thomson O'Brien MA, Freemantle N, Oxman AD et al., 2001	Continuing education meetings and workshops: effects on professional practice and health care outcomes	Cochrane Effective Practice and Organisation of Care Group specialized register, MEDLINE (from 1966), the Research and Development Resource Base in Continuing Medical Education in January 1999 and reference lists of articles 32 studies were included	Randomized trials or well designed quasi-experimental studies examining the effect of continuing education meetings (including lectures, workshops, and courses) on the clinical practice of health professionals or health care outcomes
Hogan DB, Jennett P, Freter S et al., 2001	Recommendations of the Canadian Consensus Conference on Dementia--dissemination, implementation, and evaluation of impact	A Medline search on the dissemination and evaluation of the 1989 Canadian Consensus Conference on the Assessment of Dementia (CCCAD) and other published guides for physicians on dementia care.	CCCD dissemination that has occurred to date (June, 2000) was reviewed in this paper
Amin Z, 2000	Theory and practice in continuing medical education Careful planning and evaluation of CME will improve the key measure of physician's performance and health care outcome	Two electronic databases, Medline and ERIC (Educational Research Information Clearinghouse) were searched for suitable articles published from late 70s to late 90s.	No specific criterion for article selection was mentioned.

Cantillon P, Jones R, 1999	Does continuing medical education in general practice make a difference	Medline, BIDS, ERIC, and Embase between 1990 and March 1999	(a) systematic reviews of continuing medical Education; (b) systematic reviews of postgraduate CME for general practitioners; (c) postgraduate educational interventions based on general practice. Intervention studies were included if they contained a robust evaluation, which examined either the effects of the educational event on subsequent doctor behavior or patient outcomes. Selected references from these papers were also retrieved.
Davis DA, Thomson O'Brien MA, Freemantle N et al.1999	Impact of formal continuing medical education	CME evaluation studies published between 1993 and 1999	Primary studies; more than 50% of the participants were practicing physicians; RCT of formal CME educational interventions that were didactic and/or using interactive educational techniques; objective determinations of health professional performance in the practice setting and/or determinations of healthcare outcomes

Research Basis for Instrument Development

Significance of the Problem in the U.S.

The Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (USHHS), recently released an evidenced report on the Effectiveness of Continuing Medical Education prepared by Johns Hopkins University, Evidence-based Practice Center (2007). According to the report, the literature has a low overall quality that hardly leads to consequently firm conclusions. However, the idea that CME was effective was supported by the literature. The effectiveness has been demonstrated in acquisition and retention of knowledge, attitudes, skills (56 of 69 studies, level 2), practice behaviors (61 of 105, level 3), and clinical practice outcomes (14 of 33, level 4). Only a few articles addressed internal and/or external characteristics of CME activities. Crucial factors for CME success are hard to determine due to their heterogeneity. In addition, the lack of information on reliability and validity of the instruments that have been used to assess CME effectiveness has limited the evidence.

The Office of the Director, Office of Disease Prevention (OD/ODP) of the National Institutes of Health (NIH), has developed a "white paper" on the effectiveness of CME evaluation. The conclusions were consistent with that from AHRQ report. In this white paper, Tian and colleagues conducted a comprehensive literature review of CME evaluation studies published in the years 2000 to 2006 (2007). According to Tian et al.'s review (2007), questionnaires, surveys, and scales used in these studies have not been adequate to evaluate CME outcomes related to physician knowledge, attitudes, and beliefs. Among all 32 studies reviewed, 16 of them used surveys/questionnaires specific to the clinical domains addressed by those studies without psychometric testing for validity and

reliability. Six studies adopted existing instruments in their field that had reliability and validity information. For example, one article described a study by Merckaert et al. (2005) that used the Hospital Anxiety and Depression Scale and provided validity and reliability statistics. However, the majority of studies (N=10) developed their own questionnaires, such as the diffusion and acceptability questionnaire used in Waldorff et al.'s study (2003). Six of these 16 studies (18.8%) adapted existing instruments and provided reliability and validity information. The use of self-developed questionnaires without validity and reliability information in different CME evaluation studies makes those evaluation results questionable and makes comparison across different studies impossible.

In summary, the evaluation of the CME effectiveness has been limited by the reliability and validity of the instruments that have been used currently. In the proposed study, the development of a standard questionnaire with core items on attitudes, self-efficacy, and beliefs that can be adapted for different CME programs for evaluation and comparison was conducted to fill this gap that it enables the comparison of effectiveness across different CME interventions. Comparison of these standardized results has the potential to help researchers understand factors influencing the effectiveness of different CME programs and guide future intervention design.

Theoretical Framework

In recent decades, increased attention has been given to theories and models developed within social psychology, such as the Health Belief Model and the Theory of Reasoned Action (Taylor et al., 1995, Holman et al., Lorig et al., 1992). Another major contribution was given by Bandura (1986), who developed social cognitive theory (SCT)

based on social learning theory. Both the TRA and SCT serve as the underlying theory base of the proposed instrument.

1. Theory of Planned Behavior (TPB)

One of the issues of the CME interventions is whether those changes in knowledge, beliefs, and attitudes are transferable to changes in physicians' clinical practices and further transferable to patients' health outcomes. Therefore, the critical issue is whether changes in attitudes and beliefs will be transferred to physicians' clinical practices.

The Theory of Reasoned Action (TRA) was developed by Fishbein (1967). The TRA has been used extensively in recent years to understand and predict health behaviors and to develop interventions. The TRA asserts that the most important determinant of behavior is a person's behavioral intention. Behavioral intention was defined as perceived likelihood of performing the behavior (Glanz et al. 1997). The direct determinants of behavioral intentions are attitudes toward performing the behavior and subjective norms associated with the behavior (Glanz et al. 1997). In our case, physicians' intentions to perform certain clinical practices are determined by their attitudes toward performing these clinical practices and associated subjective norms.

According to TRA, attitudes are determined by the individual's beliefs about outcomes or attributes of performing the behavior (behavioral beliefs) weighted by evaluations of those outcomes or attributes. Thus, a person who holds strong beliefs that most positively valued outcomes would result from performing a behavior will have a positive attitude toward that behavior. In our case, physicians' attitudes toward performing certain clinical practices are determined by their beliefs about the outcomes or

attributes of performing that clinical practice weighted by evaluations of those outcomes or attributes.

A person's subjective norms are determined by normative beliefs—whether important referent individuals approve or disapprove of performing the behavior, weighted by one's motivation to comply with those references. Thus, a person who believes that certain referents think he or she should perform a behavior and who is motivated to meet the expectations of those referents will hold a positive subjective norm (Glanz et al, 1997).

Ajzen and colleagues (1991) added perceived behavioral control to the TRA and proposed the Theory of Planned Behavior (TPB). A person's perceived behavioral control is determined by control beliefs—presence of absence of facilitators and barriers to behavioral performance, weighted by the perceived power—impact of each factor to facilitate or inhibit behavior. Thus, a person who holds strong control beliefs about the existence of factors that facilitate the behavior will have high perceived control over the behavior. TPB constructs of attitudes, subjective norms, and perceived behavioral control could fit into the second evaluation level of the Kirkpatrick model, i.e. learning outcomes, which involves assessment of changes in skills, knowledge, or attitudes among learners.

TPB has been used to investigate physicians' changing their clinical practices. Millstein and colleagues (1996) compared the TRA and the TPB in predicting physicians' delivery of preventive services. The findings suggested that adding perceived behavioral control to the TRA model significantly increased the variance accounted for in behavioral intention and subsequent behavior ($p < .001$). Perceived behavioral control had both

direct effects on behavior and indirect effects through social norms and behavioral intentions. In another study conducted by McDermott et al (2002), the researchers tried to investigate the relationship between physician-reported practice behavior, knowledge, and attitudes and atherosclerotic risk factor reduction in patients with peripheral arterial disease (PAD). This research suggested that physician knowledge and positive attitudes contribute to atherosclerotic risk factor reduction for patients with PAD. According to research by Montaño and colleagues (2000), physician attitude, facilitating conditions, and their interaction were identified to be significant determinants of sigmoidoscopy rate (multiple $R = 0.72$).

In a recent literature review (Jaussent, 2004), the questionnaires designed to assess the knowledge, perceptions, attitudes, beliefs, and self-efficacy also did not have a solid theoretical basis and were not based on TPB constructs. Attitudes and beliefs, two important constructs of TPB have been widely used in CME evaluation instruments and have been proven to be able to predict behavior intention of clinicians. As mentioned above, constructs of attitudes, subjective norms, and perceived behavioral control from the TPB fit well into the second evaluation level of the Kirkpatrick model (1994). Behavioral intention, however, is located between the second and third evaluation levels as it is determined by the constructs from the second level while predicting the element (behavior) in the third evaluation level. Behavioral intention questions would serve as a proxy measure of physician behavior in the third evaluation level. Further details will be provided in the following section of instruments for CME evaluation.

2. Social Cognitive Theory (SCT)

According to Social Cognitive Theory (SCT), behavior is governed by expectancies and incentives (Schwarzer et al, 1992). The likelihood that people adopt a health behavior depends on three cognitions: (a) the perception that health is threatened; (b) the expectation that behavioral change will reduce the threat (outcome expectations); (c) the values that a person places on a particular outcome (outcome expectancies), and (d) the expectancy that one is competent to change the behavior (self-efficacy) (Schwarzer et al, 1992).

The Theory of Planned Behavior provides a more detailed structure among the concepts that the proposed instrument is trying to measure, e.g. behavioral beliefs, evaluation for behavioral beliefs, perceived behavior control, and behavioral intention. However, the constructs from the Social Cognitive Theory, e.g. outcome expectancy, outcome expectation, and self-efficacy are more often measured in CME evaluation studies. Therefore, the constructs of the TPB that are parallel to the ones in the SCT could be used as proxy measures in this research.

Outcome expectations are defined as the anticipatory aspects of behavior, i.e. antecedent determinants of behavior by Bandura (1977b, 1986). A person learns that certain events are likely to occur in response to his or her behavior in a particular situation and then expect them to occur when the situation arises again. For behavior that is not habitual, people anticipate many aspects of the situation in which the behavior might be performed, develop and test strategies for dealing with the situation, and anticipate what will happen as a result of their behavior in this situation. In this way, people develop expectations about a situation and expectations for outcomes of their behavior before they

actually encounter the situation. In most cases, this anticipatory behavior reduces their anxiety and increases their ability to handle the situation. In our case, physicians may expect certain clinical outcomes to change after changing their practices based on the CME intervention. As behavioral beliefs is the parallel construct in the TPB to outcome expectations in the SCT, it could be used as a proxy measure for outcome expectations in this instrument.

Expectations are learned in four ways: (1) from previous experience in similar situations (performance attainment), (2) from observing others in similar situation (vicarious experience), (3) from hearing about similar situations from other people or social persuasion (verbal persuasion), and (4) from emotional or physical responses to behaviors (physiological state). In the context of CME, performance attainment will be physicians' personal experience in certain clinical practices, vicarious experience will be from observing peers in doing certain medical practices, and verbal persuasion will be hearing about performance changes through the CME intervention or other sources. Physiological state could be physicians' physical or emotional response after performing the newly introduced clinical practice.

Outcome expectancy (incentives, by Bandura, 1977b, 1986) is defined as the values that a person places on a particular outcome. Expectancies influence behavior according to the hedonic principle; e.g. holding everything else constant, a person will choose to perform an activity that maximizes a positive outcome or minimizes a negative outcome. In the context of CME, outcome expectancy is the value given by the physicians to the outcome resulted from changing their certain clinical practices introduced by CME intervention. In addition, assessing a person's positive expectancies early in the designed

project would be able to help identify motivators for those behaviors thus help promote health behavior changes (Glanz et al, 1997). Evaluation of behavioral belief is the parallel construct in the TPB to outcome expectancy in the SCT.

Self-efficacy is defined as the confidence a person feels about performing a particular activity, including confidence in overcoming the barriers to perform that behavior. In our case, it will be the confidence that physicians feel about performing a particular practice introduced by a CME intervention. Research by Bandura (1977a, 1978, 1982, 1986) suggested that self-efficacy is the most important prerequisite for behavioral change as it affects how much effort is investigated in the given task and what level of performance is attained (Ewart et al, 1983). According to O'Leary et al. (1985), self-efficacy plays a pivotal role in the process of behavioral change and is a primary predictor of behavioral intention (Bandura, 1977). Measurement of self-efficacy must be specific to the target behavior and to the barriers faced by the target audience and audience member's understanding and capabilities (Maibach and Murphy, 1995).

According to Glanz et al (1997), the construct of perceived behavioral control in the TPB is similar to the construct of self-efficacy by Bandura (1991), which is concerned with an individual's judgments of how well he or she can perform a behavior under various inhibiting conditions. In this research, perceived behavioral control in the TPB could be used as a proxy measure for self-efficacy in the SCT.

Studies investigating individual health behavior using self-efficacy theory have found efficacy expectations to be an important factor in the individual decision to initiate lifestyle change. However, the role of self-efficacy in maintaining change over time remains unclear.

As indicated in the recent literature reviews, the questionnaires designed to assess the attitudes and beliefs as well as self-efficacy did not have a solid theoretical basis (Jaussent et al, 2004). No CME evaluation instrument was found that measured clearly stated SCT constructs. However, a review of studies of clinician attitudes toward clinical practice guidelines revealed that two of the barriers proven to prevent physicians in following the clinical guidelines were lack of self-efficacy and lack of outcome expectancy (Cabana et al., 1999). Cabana et al's review (1999) demonstrated that lack of self-efficacy has been reported as a barrier by at least 10% of the respondents in 15 of the 19 identified surveys measuring it as a possible barrier. Likewise, lack of outcome expectancy has been reported as a barrier by at least 10% of the respondents in 7 of the 8 identified surveys measuring it as a possible barrier.

There were four studies identified in the review by Cabana et al (1999) that demonstrated both lack of self-efficacy and lack of outcome expectancy as possible barriers (Rimer et al, 1990; Grol et al, 1990; Bradley et al, 1995; CDC, 1995). In the study by Grol and colleagues (1990) the respondents' attitude to national standards were measured by seven topics, each using five point Likert scale ranging from 'strongly agree' to 'strongly disagree.' The respondent's attitude to the Nederlands Huisartsen Genootschap (NHG) as the provider of the standards were measured by three topics, each using five-point Likert scale ranging from 'strongly agree' to 'strongly disagree.' The authors suggested that research and/or interventions focused on barriers including lack of self-efficacy and lack of outcome expectancy would help improve physicians' behavior (Cabana et al., 1999).

Instruments for CME evaluation

This section summarizes the pre-existing instruments for CME evaluation addressing the variables in the second evaluation level. Among all the questionnaires reviewed with documented validity and reliability information, only three of them addressed the variables in the second evaluation level, e.g. physicians' change in knowledge, attitude, beliefs, self-efficacy, and/or skills. Each of them will be discussed individually as follows. In addition, the Diabetes Attitude Scale (DAS) will also be discussed here as it is a well validated attitude scale with reliability information provided. It has not been addressed in our recent review though (Tian et al., 2007), as the CME evaluation study using it had a physician participation rate of below 50% (i.e., 47%). The Primary Care Practitioner's Attitudes and Confidence Scale did not provide psychometric information, but the strategies in its development are worth paying attention to, thus it is also included in this section.

In a study by Sanci and colleagues (2000), the researchers developed two questionnaires for the general practitioners to rate their comfort with and their knowledge and skill of clinical processes in order to evaluate the effectiveness of an educational intervention in adolescent health care. Variables included the clinical approach to adolescents and their families and clinical processes for the substantive issues of depression, suicide risk assessment, alcohol and drug issues, eating disorders, sexual history taking, and sexual abuse. The Cronbach's alphas for comfort in clinical process and substantive issues were 0.88 and 0.93 respectively. The Cronbach's alphas for self perceived knowledge and skill in clinical process and substantive issues were 0.90 and 0.94, respectively. Short answer and multiple choice items were developed to reflect the

workshop topics and were used to assess physicians' change in knowledge. Contextual and content validity of the items was assessed through pre-testing and refined. A summary score was awarded by the course tutor without knowing the grouping information. This study demonstrated that comfort in clinical process and substantive issues, e.g. self-efficacy, was a good indicator for outcome effectiveness, and this concept should used in the CME evaluation instrument being developed.

In a study by Sanders and colleagues (2003), general practitioners reported perceived proficiency in a number of core skill domains, and their confidence in their parent consultation skills was evaluated by a self-developed Parent Consultation Skills Checklist. A seven-point Likert scale ranging from 1 (not at all confident) to 7 (very confident) was used to rate all 18 items in the scale. All the items loaded significantly on the scale after running factor analysis, indicating the checklist measured a single construct. The Cronbach's alpha for the scale was 0.97. The seven-point scale was therefore adapted as the response scale for the CME instrument being developed. However, semantic scales were used instead of Likert scales. The authors recommended that factor analysis be conducted in order to investigate the constructs measured in this CME instrument.

Jacobs et al. (2005) developed a 25-item questionnaire for participants to evaluate their self-efficacy in performing 25 surgical tasks in advanced trauma operative management. The participants' confidence in performing each surgical task was rated with a scale of 1 to 5, with a score of 1 indicating very little self-confidence and a score of 5 indicating quite a lot of self-confidence. The scale had a highest possible summate score of 125. Participants' overall confidence level was calculated by dividing each

participant's summate score by the number of items in the scale, i.e. 25. In the CME evaluation instrument, we also have a construct of self-efficacy in performing the preoperative therapy to breast cancer patients. Similar to this study, overall confidence level of physician participants could be obtained by dividing that summate score by the number of self-efficacy items.

In the Jacobs et al. study (2005), a national expert panel evaluated item content for this self-efficacy instrument. A process commonly referred to as "known groups" was used to further assess support. In this process, the instrument was administered to diverse groups of physicians including anesthesiologists, emergency department physicians, junior surgical residents, senior surgical residents, trauma fellows, attending surgeons, and expert traumatologists. Different groups of physicians were supposed to score differently on their confidence for performing the surgical procedures. For example, it was reasonable for anesthesiologists and emergency physicians to have lower self-efficacy in performing complex trauma surgery. Despite having knowledge of surgical techniques, they would not be expected to be competent in performing such procedures. However, no reliability information was provided for this self-efficacy questionnaire. According to the results of this study, the proposed study is likely to find that physician participants will have greater confidence than non-physicians in conducting the pre-operative therapy for breast cancer patients than non-physician participants. Therefore, analysis of CME evaluation findings should focus on the outcomes of specific groups of medical practitioners.

Short et al. (2006) developed and validated a scale for measuring physician readiness to manage intimate partner violence (IPV). The initial pool of items for this

survey included items adapted from existing IPV physician survey tools and items developed for the CDC and the Massachusetts Medical Society. The draft instrument included 90 proposed survey questions grouped into four major sections: (1) four background scales assessing type of previous IPV training, amount of previous IPV training in hours, perceived IPV knowledge, and perceived IPV preparation; (2) a 19-item knowledge scale with multiple choice, matching, and true/false questions; (3) IPV opinions with 54 individual questions regarding attitudes and beliefs that were scored on a seven-point Likert scale from strongly agree to strongly disagree (intentional negative wording and reversed scoring was used on some opinion items); and (4) a 13-item practice issue scale assessing self-reported behaviors (individual and office IPV practices) and policies.

The perceived preparation scale contained 11 items assessing the level of respondents' preparedness they felt for working with IPV victims. Scores and responses ranged from 1 (not prepared) to 7 (well prepared). This scale had a high internal consistency with a Cronbach's alpha of 0.96. The perceived knowledge scale included 16 items assessing the level of respondents' perceived IPV knowledge. Scores and responses ranged from 1 (nothing) to 5 (very much). This scale had a similar high internal consistency with a Cronbach's alpha of 0.96.

The final 36-item opinion scale included eighteen items from the CDC instrument, 9 items re-worded from the CDC instrument, and 9 new items. Six good-fit scales (preparation, legal requirements, work place issues, self-efficacy, alcohol/drugs, and victim understanding) with 31 items were identified in this section with a Cronbach's alpha of 0.65. Construct validity was demonstrated by the significant correlations

between perceived knowledge score and the amount of previous training and perceived preparation scales. The scales moved in the same direction as they measured different aspects of a physician's preparedness to manage IPV.

We can see from this study that although no health behavior theory was mentioned as the basis for developing the scale, attitudes and beliefs—the concepts of the TPB—were actually used here as indicators to evaluate the program effectiveness. The response scales for items addressing these concepts were also seven-point Likert scales from strongly agree to strongly disagree, and intentional negative wording and reversed scoring was used on some opinion items. Seven-point semantic scaling appeared to be the appropriate choice of the response scale for the CME instrument being developed given the current literature and the scale development theories (Tourangeau et al., 2000). This decision was justified later with further citations in the methodology Chapter.

Sharp and Lipsky's research (2002) used the third version of the Diabetes Attitude Scale (DAS-3, Anderson et al., 1998) to measure health care providers' attitudes towards diabetes and towards treatment of diabetes. DAS-3 is a 33-item experience-based self-report instrument modified in 1998 from an earlier version (Anderson et al., 1989). Each item reflects a belief about diabetes or the treatment of diabetes. Attitude subscales were comprised by scoring all the similar beliefs together: (1) "Special training" reflects an attitude supporting the benefit of special training in communication, patient education, and counseling for health care providers who work with diabetes; (2) "Seriousness of type 2" reflects the attitude that type 2 diabetes is a serious disease; (3) "Tight control" reflects the attitude that tight control of serum glucose can prevent complications; (4) "Psychosocial impact" reflects the attitude that diabetes has a negative psychosocial

impact for the patient; and (5) “Patient autonomy” reflects the attitude that encouraging the patient to make daily decisions in managing diabetes is valued.

Content validity was determined by diabetes experts at the University of Michigan Diabetes Research and Training Center. Responses were based on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The administration time was 10 to 15 minutes. Higher scores represented more positive attitudes in each area of care. The five DAS-3 subscales had Cronbach’s alpha reliability coefficients ranging from .65 to .80. The subscales were moderately correlated with coefficients ranging from .27 to .63.

In another study conducted by Henderson et al. (2005), the attitudes and confidence of primary care providers (PCPs) were evaluated as proxy measures of clinical performance. The study demonstrated that a high level of physician confidence in performing a medical procedure or psychosocial task was closely correlated with their actual performance of the procedure (Bernard et al., 1999; Davis et al., 1997; Smith et al., 1994, 1998, 2000; Wickstrom et al., 2000a, 2000b). Confidence in the areas of medical evaluations (4 items), psychiatric diagnosis (9 items), prescribing psychotropic medications (8 items), providing counseling for psychiatric disorders (8 items), treating patients with a history of violence or victims of violence (4 items), and traditional healing (2 items) was evaluated by the survey. A 6-point Likert scale ranging from (1 = not at all confident, 2 = slightly confident, 3 = somewhat confident, 4 = confident, 5 = very confident, 6 = extremely confident) was used to measure PCP confidence to perform medical and psychiatric procedures. However, no information was provided showing that this instrument was validated.

Henderson and colleagues (2005) cited Smith et al.'s study (1998) and Wickstrom et al.'s study (2000a, 2000b) to support their evaluation of attitudes and confidence of primary care providers (PCPs) as proxy measures of clinical performance. According to Smith et al. (1998) and Wickstrom et al. (2000a, 2000b), confidence does not necessarily reflect one's competence, but self-belief in one's ability to perform certain procedures may approximate competence. Smith et al. (1998) used 38-item questionnaires to examine self-confidence in five skill areas and found that trained residents expressed higher self-confidence in all five areas of psychosocial skill and anticipated more positive outcomes for emotional sensitivity ($p = 0.05$), managing somatization ($p = 0.03$), and facilitating patient communication ($p = 0.02$). Trained residents were also more strongly committed to being emotionally sensitive ($p = 0.055$) and managing somatization ($p = 0.056$) compared with the untrained residents (Smith et al., 1998; Wickstrom et al., 2000a, 2000b). Their increased confidence may merely reflect changes in attitudes.

Like in the previous research, no theoretical basis was clearly stated for developing the instrument in the study, but the indicators being used—attitudes, self-efficacy (perceived behavioral control, self-confidence in clinical skills)—were consistent with those from TPB and SCT. The study indicated that self-efficacy (perceived behavioral control) could serve as the predictor of future clinical behaviors.

In summary, these instruments provide guidance to the conduct of the proposed study. Although lack of clear stated theoretical basis, the concepts of attitudes, beliefs and self-efficacy (perceived behavioral control) are widely used in evaluation instruments; self-efficacy also has been proven to predict future clinical behaviors. In addition, as far as response scale, 7-point Likert scaling has been used for these two concepts and would

be the appropriate choice for the proposed instrument. Cronbach's alpha is used as indicator of reliability for all the instruments being reviewed, and it will be the one for our instrument, too.

Summary

In summary, the review of the literature revealed that the heterogeneity of CME programs and evaluation efforts limits the determination of the key factors for their success. The evidence was limited by the reliability and validity of the instruments that have been used to assess CME effectiveness (AHRQ, 2007). In addition, current CME evaluation instruments addressing the variables in the second evaluation level usually lack a theory base. They usually lack of validity and reliability information as well (Tian et al, 2007). Only after we get strong measures of attitude and behavioral intention change from level two, should we focus further on the physician and patients outcome level three and four (Kirkpatrick, 1994). Thus, the development of a valid and reliable instrument addressing physicians' changes in the concepts listed in the second evaluation level was deemed useful. In addition, most of the constructs being evaluated in the developed instrument (i.e. belief, attitude, perceived behavioral control and subjective norms) belong to the second evaluation level of Kirkpatrick's Outcome Evaluation Model (1994). Behavioral intention, however, is located between the second and third evaluation levels and is determined by the constructs from the second level while predicting the element (behavior) in the third level.

According to Lorriman (1997), factors that contribute to professional competence are attitudes, knowledge, and skills. Based on the literature of the modified Kirkpatrick's

model, the Theory of Reasoned Action, Social Cognitive Theory, and the review of validated instruments that have been used in CME evaluation studies, the constructs of beliefs, subjective norms, perceived behavioral control, and behavioral intention should be included in the scale to evaluate the intervention outcomes in the second evaluation level.

This literature suggested that the constructs to be evaluated in the standard instrument should be (1) behavioral beliefs for performing a certain clinical practice, (2) attitudes towards performing a certain clinical practice; (3) subjective norms of performing a specific clinical practice; (4) perceived behavioral control (self-efficacy) in performing a specific clinical practice; and (5) behavioral intention to perform a specific clinical practice. Such items could be modified and applied to various CME programs in different clinical dimensions. This would improve the validity of results for those interventions as well as facilitate the comparison of the effectiveness across different CME interventions. This in turn, will help the researchers understand the characteristics of effective CME programs and apply them when planning and implementing future CME programs.

Chapter 3: Methodology

The purpose of the research was to develop a standardized, theory-based, valid and reliable CME evaluation instrument for clinicians assessing the constructs addressed in the second evaluation level of Kirkpatrick's model. This chapter presents the methodology for developing the instrument, the selection of measurement format, the selection of experts, the methods used to conduct the pilot test and the proposed study, and the procedures for data analyses and finalizing the scale. Scale development is a multi-step procedure, and this chapter includes the following three phases: (1) phase I, scale development; (2) phase II, scale validation; and (3) phase III, data collection and analyses. Expert judgment was employed to examine the instrument for face validity. Reliability and factor analysis were employed to examine evidence for evaluating the construct validity. The rationale for using these theories and methods are presented.

Phase I: Scale Development

Scale development includes several steps (Babbie, 2000). They are: 1) develop template items addressing variables in the second evaluation level based on the TPB, i.e. attitude, behavioral belief, subjective norm, perceived behavioral control and behavioral intention; 2) determine the format for measurement as 7-point semantic scale; 3) develop items specialized to the conference of Preoperative Therapy in Invasive Breast Cancer; 4) have the template reviewed by experts (CME committee members, University of Maryland faculty members); 5) have the initial item pool reviewed by experts (CME committee members, conference organizers and conference instructors) to examine the face validity; 6) finalize the initial pool of evaluation items; 7) conduct cognitive testing

with the target population; and 8) pilot test the instrument with the target population and make revisions as needed in order to improve the internal validity.

Evaluation Instrument Template

Skills and knowledge are extremely content specific and less likely to be modified from the general items being developed. The aforementioned concepts (i.e. behavioral beliefs, attitudes, subjective norms, perceived behavioral control and behavioral intention), however, can be easily modified according to specific clinical dimensions from more general items. As a result, knowledge and skills were not addressed in this standardized instrument. See Figure 3-1 for the instrument template.

The instrument was to be administered to the participating physicians of the meeting of Preoperative Therapy in Invasive Breast Cancer: Reviewing the State of the Science and Exploring New Research Directions, which was held March 26 and 27, 2007, in the Natcher Conference Center, National Institutes of Health in Bethesda, Maryland. The clinical domain of items was adapted to the one addressed in this conference, i.e. preoperative therapy in invasive breast cancer. The meeting purpose and the conference objectives were used to operationalize adaptable measures for CME activities designed to address physician practices. Figure 3-2 summarizes the meeting purpose, and Figure 3-3 presents the course objectives. See Appendix G for the questionnaire adapted to this conference.

Figure 3-1. Instrument Template

Demographics

Please circle your specialty			
Please circle your affiliation	Academia	Government	Industry
	Community Practice	Other(Please specify) _____	
Please indicate number of years in practice:			
Your date of birth (mm/dd/yy):			
Your initials:			
Are you seeking CME credits?		YES	NO
Your gender:		FEMALE	MALE

Behavioral Beliefs

[Medical procedure] will *lead to a lower mortality rate* of the [type of] patients.	Unlikely (1)—likely (7)
[Medical procedure] will *improve the [medical index]* of [type of] patients.	
[Medical procedure] will have *fewer side effects* for [type of] patients.	
[Medical procedure] will *reduce the overall medical costs* for [type of] patients.	

Attitudes

The practice of pre-operative systemic chemotherapy is…								
Not credible	1	2	3	4	5	6	7	Credible
Unsafe	1	2	3	4	5	6	7	Safe
Harmful	1	2	3	4	5	6	7	Beneficial
Ineffective	1	2	3	4	5	6	7	Effective
Frustrating	1	2	3	4	5	6	7	Satisfying
Impractical	1	2	3	4	5	6	7	Useful
Hard	1	2	3	4	5	6	7	Easy

Subjective Norms

My colleagues think I should share information about	

[Medical procedure] with the [type of] patients.	
My colleagues think I should share knowledge of [Medical procedure] with physicians who do not attend the conference.	
My colleagues think I should recommend [Medical procedure] to the [type of] patients	
My colleagues think I should refer the [type of] patients to the trails of [Medical procedure]	

Behavioral Intention

I intend to	Unlikely (1)—likely (7)
share information about [medical procedure] with [type of] patients.	
share knowledge of [medical procedure] with physicians who do not attend the conference.	
review the literature about [medical procedure].	
apply knowledge of [medical procedure] in developing research studies.	
evaluate/assess [type of] patients' eligibility of receiving [medical procedure] in [specific instance].	
recommend [medical procedure] to [type of] patients in [specific instance].	
provide [medical procedure] to [type of] patients in [specific instance].	
refer [type of] patients to appropriate trials in [medical procedure].	

Perceived Behavioral Control (Self-Efficacy)

Rate your confidence level in	
sharing information about [medical procedure] with [type of] patients.	Unconfident (1) --- Confident (7)
sharing knowledge of [medical procedure] with physicians who do not attend the conference.	
evaluating/assessing [type of] patients' eligibility of receiving [medical procedure] in [specific instance].	
recommending [medical procedure] to [type of] patients in [specific instance].	
providing [medical procedure] to [type of] patients	

in [specific instance].	
referring [type of] patients to appropriate trials in [medical procedure].	
applying knowledge of [medical procedure] in developing research studies.	
evaluating [medical procedure] papers critically when they appear in the literature.	

Figure 3-2 Purpose of the Meeting: Preoperative Therapy in Invasive Breast

Meeting Purpose

According to the conference information released by NIH "Preoperative therapy is increasingly being administered to women with breast cancer. Controversies exist, however, regarding optimal approaches." This NCI-sponsored conference will seek to determine the state of the science regarding clinical use of preoperative therapy in breast cancer, as well as identify future research agendas. The conference will seek to answer the following questions:

- What is established in the field of preoperative therapy for invasive breast cancer?
- How should what is known be properly applied?
- How should preoperative therapy be incorporated into research initiatives?

Leading breast cancer physicians will present the state of the science and engage in a panel discussion. Audience participation will be encouraged. (http://ctep.cancer.gov/bcmeeting/)

Figure 3-3 Educational Objectives of the Meeting

Educational Objectives

According to the conference announcement, participants who attend should be able to:

- Explain the state of the science related to key issues in breast cancer preoperative therapy, and apply this knowledge in the treatment of patients in clinical practice and research, including patient evaluation, treatment selection, response monitoring, and locoregional management.
- List or discuss advantages or benefits of systemic therapy before surgery.
- List or discuss advantages or benefits of radiation therapy before surgery.
- Apply knowledge of preoperative therapy in developing research studies or critically evaluating them when they appear in the literature.
- Evaluate breast cancer patients before, during, and after preoperative therapy using appropriate endpoints and prognostic biomarkers to guide treatment decisions.
- Describe special issues related to locally advanced breast cancer (LABC) and inflammatory breast cancer (IBC) in the preoperative setting, and apply this knowledge in clinical practice and/or research.
- Refer patients to appropriate trials in breast cancer preoperative therapy. (http://ctep.cancer.gov/bcmeeting/)

Item Development

As mentioned in the last chapter, five constructs based on the Theory of Planned Behavior were addressed in this instrument (Figure 3-4), namely:

(1) Behavioral beliefs of performing a certain clinical practice;

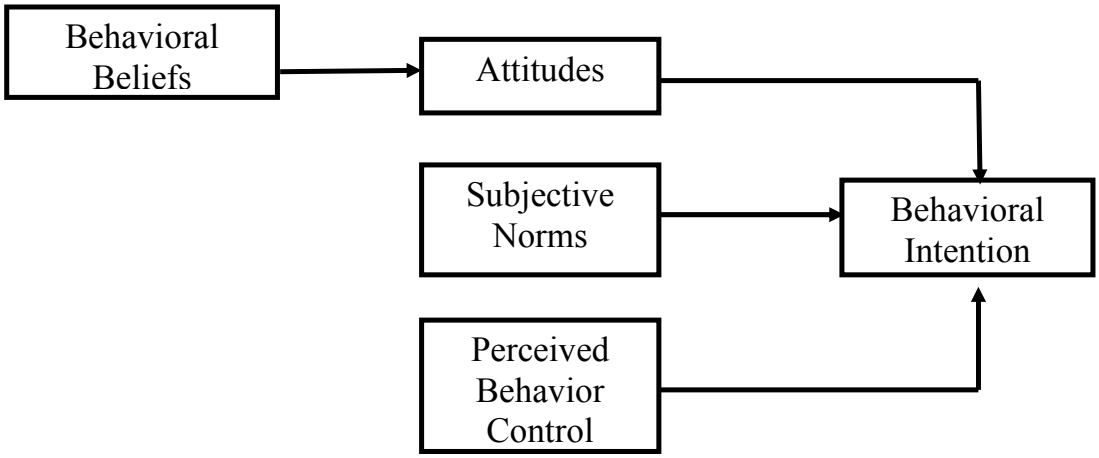
(2) Attitudes towards performing a certain clinical practice;

(3) Subjective norms of performing a certain clinical practice;

(4) Perceived behavioral control for performing a specific clinical practice; and

(5) Behavioral intention to perform a specific clinical practice.

Figure 3-4 Theoreticl Framwork



As the items in the instrument were clinical domain specific, and there was no existing validated instrument in this clinical area (i.e. breast cancer preoperative therapy), new items were developed for every single construct (i.e., behavioral belief/outcome expectation, attitude, subjective norms, perceived behavioral control/self-efficacy and behavioral intention). In addition, in order to assure that the items could be easily understood, several considerations were made. First, clear and unambiguous languages were used in writing the items. All abbreviations were defined for the physicians. A double-sided, single page of close-ended items was

developed to guarantee that the survey could be completed in a minimum amount of time.

Format of Measurement

Seven-point semantic scales were used for each item in the proposed instrument. Ajzen and Fishbein (1980) have clearly defined the measurement of model components in the TPB and causal relationships among these concepts. They suggested a person's behavioral beliefs about the likelihood that performing the behavior will result in certain outcomes should be measured on bipolar "unlikely"-"likely" scale. Attitudes towards performing the behavior were measured on seven bipolar semantic scales.

According to Tourangeau et al. (2000, p244), "the inclusion for a midpoint seemed to reduce positivity bias, mostly by drawing responses from the positive end of the scale to the midpoint. The verbal label may have made the meaning of the numeric midpoint clearer or it may have merely made that response option more salient." Instrument development literature has suggested that all the response categories for the questions in the proposed instrument, especially the midpoint, be given a verbal label in order to improve the reliability and reduce positivity bias (Tourangeau et al., 2000). The experts reviewing the survey were asked what type of response categories would be easier for physicians to fill out: verbal labels, numbers, or blank boxes (Appendix A, B, C). All the experts preferred the numeric response format (Appendix B). Therefore, numbers instead of verbal labels were used for the responses to the scale items.

In addition, Tourangeau et al. (2000) argued that the ratings tended to fall on the positive half of the scale regardless of labels, which is in line with the positivity bias. Negative numbers convey a different meaning from labels that range from 0 and up.

Rather than being a logical complement of a high end, negative numbers imply that the end of the scale is the polar opposite. This implication may be misleading with a unipolar dimension (Tourangeau et al., 2000). Given the aforementioned the reason, positive numbering (scored from 1 to 7) was used in the responses for all the scale items in order to avoid positivity bias.

Expert Feedback

Expert feedback was sought during the construction of preliminary items. NIH CME committee members, faculty members of the University of Maryland, and conference organizers and instructors were asked to provide feedback while the instrument was under construction. The experts were looking at the adaptable instrument as well as the instrument for the conference. The adequacy of the items for each subscale was discussed and additional items were suggested. In addition, the weaknesses in the following areas were discussed, e.g. content, representation of the constructs, use of unidimensional statements, unambiguous wording, readability, and clear instructions were discussed until consensus was reached (Flower, 1995).

Meeting organizers and instructors were asked to review items to ensure that the scope of the information delivered in the conference was represented in order to enhance the content validity. In addition, a professor with expertise in measurement and quantitative methods examined the psychometric scales for dimensionality and optimal category usage related issues. Questions were revised accordingly, and the initial instruments were constructed and formatted for the cognitive testing and pilot study (Appendix D).

Cognitive Testing

Four physicians from the target population were referred by the meeting organizer to conduct the cognitive testing for the instrument. Cognitive Testing Instruction can

be found in Appendix E. The clinicians were asked to walk through the instrument. They were asked to read the questions aloud to themselves and talk out loud about their reactions if any of the questions were difficult to understand, were hard to answer, or did not make sense. They were also asked to answer the questions one by one and tell what responses they selected and why they selected them. Revisions to the instrument were then made accordingly.

Phase II Scale Validation

Pilot Test

In social science research, pilot studies are used in two different ways. First, they can refer to "small scale version[s], or trial run[s], done in preparation for the major study" (Polit et al., 2001)—so-called feasibility studies. Second, they can be used to pre-test or 'try out' a particular research instrument (Baker, 1994). The purpose of doing a pilot study for this research would be the second one, to pre-test the CME evaluation instrument to improve its internal validity, for example, if the instrument developed is inappropriate or too complicated. The procedures of conducting the pilot study listed below were adapted from Peat et al.'s study (2002).

First, CME evaluation instruments with items that result from the cognitive testing and expert review were administered to the pilot subjects, four medical oncologists. Facsimiles were used to collect the completed instruments since all the participants were not local. Participants were asked for feedback to identify ambiguities and difficult questions. Time taken to complete the instrument was recorded, and its reasonability was decided. All unnecessary, difficult or ambiguous questions were discarded or revised. Next, the pilot testing was used to make sure that replies could be interpreted in terms of the information that was required. Finally, the

instrument was checked to see if all questions in the instrument were answered, and those questions that were not answered as expected were revised.

Phase III Data Collection and Analysis

The final paper-pencil instruments were administered at the two-day conference. SPSS 14.0 was used to perform analyses for psychometric evidence on the final scales. Item analysis, exploratory factor analysis, and estimations of reliability and validity were performed with the pre-testing data. T he data analysis section below provides the statistical procedures in further detail.

Participants

There were 269 on-site participants for the conference of Preoperative Therapy in Invasive Breast Cancer. The target audiences were breast cancer physicians (medical oncologists, radiation oncologists, radiologists, pathologists, surgeons, and others) as well as general interventional radiologists and surgeons.

Instrument Administration and Data Collection

After obtaining IRB approval (Appendix F), consent forms along with the traditional (paper-pencil) self-administered instrument were placed in the conference registration package. Clinicians' attending the NCI conference received the registration package at the registration desk and were informed about the instrument. Requests for filling out the instruments before the conference started were announced three times both verbally and visually to encourage more clinicians to participate. Clinicians put their completed questionnaires into the box sat on the registration desk. Instruments were collected at the end of both days of the conference. The pre-test data collected were used to develop the CME instrument. The consent forms were

administered along with the pre-test instruments. They described the nature of the evaluation survey, described the cooperation requested from the participants, and assured privacy and confidentiality for the participants.

Items for all the concepts were rated on a 7-point semantic scale. Responses to behavioral beliefs (outcome expectation) questions ranged from 1 "unlikely" to 7 "likely." Responses to attitudes questions ranged from 1 "unlikely" to 7 "likely." Subjective norm questions ranged from 1 "unlikely" to 7 "likely." Responses to perceived behavioral control (self-efficacy) questions ranged from 1 "unconfident" to 7 "confident." Behavioral intention questions ranged from 1 "unlikely" to 7 "likely." A sample of the final instrument is provided in Appendix G.

All the participating physicians attending the conference who agreed to participate were included in the sample. The instrument was attached to the required NIH CME evaluation form (e.g. name, birthday, professional degree, NIH badge number, phone, email, organization, institute/center, department/branch, address, rating of objectives and activity). However, private information required by NIH was not input in the study dataset. Only the instrument data were documented in the final study dataset. Whether the participant was registering for CME credits or not was also included in the questionnaire. The pre-testing data were used for item analysis, exploratory factor analysis, and estimations of reliability and validity.

Data Analysis Plan to Answer Research Questions

Statistical procedures were presented in this section to answer each of the research questions. The research questions include:

1. Will a thorough content validation process satisfy the needs for instrument validity?

2. Will a psychometric examination of the draft instrument reveal any unexpected measurement subscales?
3. Can an instrument for this purpose be developed with subscales consistent with the theoretical domains?
4. Can an instrument be developed with acceptable levels of reliability for each of the necessary subscales?
5. Will a thorough instrument development process result in an instrument that is appropriate for evaluation of CME?

Exploratory factor analysis was conducted with SPSS 14.0 software to investigate whether different variables in the scale loaded on specific factors/theoretical domains they were assumed to and whether there were unexpected measurement subscales. Item analysis was conducted for each subscale in order to guarantee that appropriate items are included in these subscales. Reliability analysis was performed using SPSS 14.0 software for each of the necessary subscales. Correlations between variables were calculated in order to examine the construct validity. Divergent validity and convergent validity were not conducted given the single instrument administration method.

1. **Will a thorough content validation process satisfy the needs for instrument validity?**

The initial draft of questionnaire included three separate sections. Section one was the instruction for completing the questionnaire. Section two included demographic questions such as age, affiliation, and specialty. Section three included 32 items addressing the constructs in the second evaluation level. The draft instrument could be found in Appendix D.

Cognitive testing, expert review, and pilot testing procedures for the instrument development were described above in detail. Cognitive testing was conducted with four physicians from the target population. The item modification suggestions were then collected and sent to the experts for review. The instrument was then revised based on experts' decisions in response to the results of the cognitive testing. After that, four other clinicians from the target population were selected to pilot test the instrument and then gave their comments. All the comments were then sent to the experts for review. T he instrument was finalized with the experts' suggestions in response to the pilot testing results.

This thorough content validation process including cognitive testing, experts' review and pilot testing satisfied the needs for a process to assure instrument validity. The instrument was shown to have a good content validity as each iteration required fewer adjustments since there were fewer recommendations from the target audience and expert reviewers.

2. Will a psychometric examination of the draft instrument reveal any unexpected measurement subscales?

3. Can an instrument for this purpose be developed with subscales in each of the theoretical domains?

Sample size

Sample size was checked to ensure that it met the sample size/variable ratio criteria suggested in the literature. Nurally (1978) suggested the need for 10 participants per item. Gorsuch (1983) suggested that the ratio be 5 to 1. These were the two main standards considered in this research. As there were 32 items in the final instrument, the 10 to 1 ratio was not met given 134 valid cases. The item to

sample ratio was 6.09 to 1 in this study. This ratio meets Gorsuch's criteria.

Exploratory factor analysis was used to reduce the items in the proposed instrument to several subscales. The items in each individual subscale were then examined to see if they were from the same theoretical domain as proposed or if they were unexpected measurement subscales.

Factor Analysis

According to Hair (1998), unidimensionality of items, e.g. strong association with each other and representing a single concept, is the underlying assumption and essential requirement for creating a summated scale. All of the five subscales in this CME instrument were intended to be summative scales. By determining the number of factors and the loadings of each variable on the factor(s), factor analysis technique helped make an empirical assessment of the dimensionality of all items in this instrument. The test of unidimensionality is that each summated scale should consist of items loading highly on a single factor (Hair, 1998).

Guadagnoli and Velicer (1988) suggested that as the number of variables loading strongly per factor increases, the number of participants actually decreases. The specific rules are as follows: 1) any factor with at least 3 loadings above .80 in absolute value should be considered reliable; 2) factors with 4 or more loadings above .60 in absolute value should be considered reliable, regardless of sample size; 3) factors with about 10 or more loadings (around 0.4 in absolute value) should be considered reliable, as long as sample size is greater than about 150, 4) factors with only a few loadings should not be interpreted unless sample size is at least 300. As our sample size was 134 physicians for the exploratory factor analysis, factors with only a few loadings were interpreted with caution.

Exploratory Factor Analysis (EFA)

Exploratory factor analysis was conducted on the pre-test data of the 134 participating physicians in order to investigate whether the items in the instrument loaded on the four theoretical domains as they were proposed to, e.g. if the items in the instrument can be grouped and reduced to four conceptual subscales measuring one of each proposed concepts, namely, behavioral belief, attitude, subjective norms, perceived behavioral control, and behavioral intention.

Tests were undertaken to determine the factorability of the correlation matrix. The assumption here is that variables should significantly correlate with each other because then they are measuring the same construct. Sample adequacy is tested by the Kaiser-Meyer-Olkin (KMO) statistics. KMO served as a qualitative index of the strength of relations among variables, based on correlation and partial correlation. Bartlett's Test of Sphericity test was also conducted to check if the correlation matrix was an identity matrix. Since the null hypotheses assume that the intercorrelation matrix comes from a population in which the variables are noncollinear (e.g. an identity matrix), nonsignificant result of this statistic permitted the factor analysis.

PCA was conducted to perform an initial extraction and to determine how many factors should be retained on the basis of multiple criteria, e.g., Kaiser's rule (extract all components with eigenvalues greater than or equal to one), Cattell's scree plot (look for big breaks in the scree plot), and Cattell-Nelson-Gorsuch's "objective" scree. Convergence among the application of multiple criteria was sought. Literature suggests that the popular Kaiser's rule will often lead to overestimate the number of component of factors that underlie the data (e.g., Cattell & Jaspers, 1967; Fava & Velicer, 1992; Hakstian, Roger, & Cattell, 1982; Lee & Comrey, 1979; Zwick & Velicer, 1986).

Principal Axis Factoring (PAF) was then undertaken, retaining as many factors as the PCA. PAF is a common factor analysis solution that places on the diagonal of the correlation matrix squared multiple correlation (R^2) of each item with all of the other items included in factor analysis. Nunnally and Bernstein (1994) argued that a PAF solution that retains as many factors as PCA will provide a better estimate of the correlations because PCA does not separate out error measurement (Pett, Lackey, & Sullivan, 2003, p.110). Widaman (1993) noted, if a researcher wishes to obtain parameters reflecting latent constructs or factors, principal component analysis should not be used. However, PAF has the disadvantage that it can generate negative eigenvalues that are meaningless. Initial and extracted communalities generated in PCA and PAF were compared even though for most datasets, PCA and PAF will lead to similar substantive conclusions (Wilkinson, Blank, and Gruber, 1996).

Oblique rotation was conducted at this step to derive a simple structure composed of factors that are easy to interpret. Direct oblimin rotation maximizes difference between the high and low loadings on a particular factor, thus simplifying the columns of the unrotated factor-loading. The assumption is that the underlying factors are correlated with each other. Since all the concepts to be measured in this CME evaluation instrument were assumed to be correlated, oblique underlying rotation was the appropriate strategy used.

Comrey and Lee (1992) generated the following guidelines for item-to-factor loadings in factor structure matrix of oblique solutions to help determine if an item should be included in defining the factor: loadings greater than .71 are excellent; loadings greater than .63 are very good; loadings greater than .55 are good; and loadings greater than .45 are fair. According to this guideline, higher factor loadings reflect a higher degree of overlapping true variance between the item and the factor;

greater number of substantial loadings on the factor indicates that it is easier to isolate what the factor potentially represents. Practically speaking, loadings less than .30 are "weak" loadings. Since loadings less than 0.55 were considered as "fair," items with loadings less than .50 were dropped from the instrument. This rigid criterion provided higher validity for the research.

Items that load significantly on more than one factor should be placed with the factor that is conceptually most closely related to (Pett, Lackey, & Sullivan, 2003). In addition, further reliability analysis was conducted for each of the factors/ subscales on which these items loaded. Cronbach's alpha values were used to evaluate the internal consistency of each factors/subscales with and without the overlapping loaded items and decide where to best place these items.

4. Can an instrument be developed with acceptable levels of reliability for each of the necessary subscales?

In order to guarantee that appropriate items were included in the revealed subscales, item analyses were conducted for each subscale. Item to total correlations and inter-item correlations were used to conduct the item analysis. Corrected-item-total correlations, alpha-if-item-deleted, and factor analysis were used to examine the internal consistency reliability of the subscales. Cronbach's alpha was also examined.

Item to Total Correlations

SPSS 14.0 software was used to generate the item-total correlation and corrected item-total correlations to make judgments about which items to retain for the final subscales. The developed multi-item scale were a summate scale for attitude, subjective norms, beliefs, perceived behavioral control, and behavioral intentions.

The total score of each subscale was the scale's "total." Corrected item-total correlations and alpha-if-item-deleted for each item in the scales were examined in order to decide whether each item should be kept or dropped.

Alpha-if-item-deleted for the items were compared with coefficient alpha for the entire subscale. There is no fixed rule for how low the correlation should be to drop the item. However, if the item has much lower (less than 10%) corrected item-total correlation than that of other items and its value of alpha-if-item-deleted is higher than the coefficient alpha for the entire subscale, it was considered for elimination. A much lower corrected item-total correlation means a certain item is not as good as other items in the scale, i.e. it is not as closely associated with the rest of the scale as the other items are. Consideration may be given to either remove or revise the item.

Inter-Item Correlations

Inter-item correlations can be obtained from computing the correlations for each pair of items. SPSS 14.0 software was used to generate the inter-item correlation matrixes for the five subscales. Average inter-item correlations use all of the items in each subscale to measure the specific constructs for that scale.

Usually, a correlation value of 0.25 to 0.3 could be considered as the existence of the correlation between two items; a correlation value of 0.5 could be considered as good correlation; a correlation value below 0.2 would be considered as divergent correlation (Trochim, 2001). Items with a divergent correlation from their subscales were removed from instrument.

Reliability

The unidimensional scaling method assumes that the concepts being assessed

are one-dimensional in nature. According to Trochim (2001), the single measurement instrument administered to the participants on one occasion would be used in estimating the internal consistency, e.g. reliability estimation. The reliability of the instrument was judged by how consistent the results are for different items reflecting the same construct within the instrument. Cronbach's alpha was used in this study for the reliability estimate.

Internal Consistency Reliability

Three common analyses are usually used to improve internal consistency, e.g. corrected-item-total correlations, alpha-if-item-deleted, and factor analysis (Trochim, 2001). These analyses were performed to help modify the composition of items in each subscale in order to improve their internal consistency.

Cronbach's Alpha

According to Trochim (2001), Cronbach's alpha is the specific method of estimating the reliability that it could be thought of as analogous to the average of all possible split-half correlations. It has been agreed in general that the accepted lower limit for Cronbach's alpha is 0.7, while it may decrease to 0.6 in exploratory research (Hair, 1998). In our research, Cronbach's alphas of 0.6 and above for the scale correlation matrixes were acceptable values.

According to Hair (1998), Cronbach's alpha is positively related to the number of items in the scale. Increasing the number of items in the scale will increase the reliability value even with the same degree of inter-correlation.

5. Will a thorough instrument development process result in an instrument that is appropriate for evaluation of CME?

Two separate sets of work have been done to answer this question. The first set of work—whether the CME instrument developed can evaluate the CME conference for pre-operative therapy—was answered by the results of cognitive testing, expert opinion, pilot testing, and the data analyses for the conference presented in the aforementioned sections answering all the previous four research questions.

The second part of the question was interpreted as whether the questions themselves with their corresponding constructs can be adopted to evaluate other CME conference by revising the content. A final adaptable instrument template and a step-by-step guide for adapting the instrument for other CME conferences were developed as reference for future CME evaluation researchers. Since the instrument has not been applied to other CME conferences, this part of the question can only be answered by future research.

Human Subjects Concerns

Both the evaluation instrument and the consent form were submitted to the University of Maryland Institutional Review Board for approval before administering them to the conference participants (Appendix G). Information was collected with confidentiality, i.e. hand written numbers were used as the identifier that participants' information could not be identified by the principal investigator (PI) either directly or through identifiers. Data were reported in aggregate form thus individual identification was not tied to data analysis and reporting.

Completed surveys were collected by conference instructors and placed in a sealed envelope. Surveys were removed only by the NIH CME committee members or UMD researchers. Data from the survey were coded for easy analyzing, interpreting, and reporting. Surveys were kept at the Public Health Informatics Research Laboratory at the University of Maryland in a locked filing cabinet. Only

CME committee members and project researchers at the University of Maryland had the access to them. Surveys were returned to NIH CME office and were shredded upon completion of the research.

There were no physical, social, or legal risks of any kind to the participants. Response to the questions in the scale would not cause discomfort or anxiety among participants. The project was designed to help develop the evaluation instrument as well as assess the effectiveness of the conference intervention. Individual participants' attitudes, beliefs, perceived behavioral controls, subjective norms, and behavioral intentions were not the focus of this study but only used to evaluate the intervention.

The participants might not have benefited directly from participating in the project and filling out the survey. However, the information collected from this project could help the development of a valid, reliable, and adaptable evaluation instrument for NIH to use in its future conferences. Hopefully, future NIH CME conference instructions and evaluations will benefit from this instrument. As the scale questions were attached to the official NIH CME evaluation form, participants were encouraged to fill out and return the survey to the fullest extent possible.

Delimitations

The following delimitations or constraints should be considered when interpreting the study findings. One, participants were physicians voluntarily registered for the NCI-sponsored CME conference of Preoperative Therapy in Invasive Breast Cancer in Natcher Conference Center, National Institutes of Health, Bethesda, Maryland. Two, data were collected in the conference on March 26 and 27, 2007, only from physicians attending in person. Three, data were collected in a way that the

proposed instrument was attached to the required NIH CME evaluation form. Finally, the data were collected in a paper-pencil format.

Summary

This chapter proposed and described plans for the development and testing of the instrument measuring attitudes, behavioral beliefs, subjective norms, perceived behavioral control, and behavioral intention among physicians attending a CME conference. The analyses included developing, validating and finalizing the instrument. The analyses were based on the data collected from the NIH CME conference. Exploratory factor analyses were conducted to develop the subscales. Reliability and validity for each of the subscale were examined. The reliable and valid instrument was developed after the analyses.

Chapter 4: Results

The purpose of this section is to provide statistical evidence on the feasibility of developing a valid and reliable instrument that measures behavioral intentions and its determinants among practicing physicians taking CME courses. Description of the instrument development and refinement process is presented first (to answer the first research question), followed by a description of the sample population, then the results of the validity and reliability analyses to answer the remaining four research questions.

Data were collected at a CME-related conference Preoperative Therapy in Invasive Breast Cancer: Reviewing the State of the Science and Exploring New Research Directions. The conference was held March 26 and 27, 2007 in the Natcher Conference Center, National Institutes of Health in Bethesda, Maryland. The purposes of this NCI-sponsored conference were to determine the state of the science regarding clinical use of preoperative therapy in breast cancer as well as identify future research agendas. Exploratory factor analysis was used to examine the underlying structure of the data and to reduce the items in the proposed instrument to six subscales. Factor loadings of the items in each individual subscale were examined to see if they were from the same theoretical domain as proposed or if they measured unexpected subscales. The consistency between the a priori subscales and the factors that emerged served as evidence for content validity of the instrument.

Item to total correlations and inter-item correlations were conducted for each subscale in order to guarantee the inclusion of appropriate items in the revealed subscales. Corrected-item-total correlations and alpha-if-item deleted were used to examine the optimal internal consistency reliability of the subscales as measured by Cronbach's alpha statistics for each subscale.

Content Validity

The results from cognitive testing, expert review, and pilot testing answered question 1 regarding use of a thorough content validation process to facilitate instrument validity. Research questions 2 and 3 were examined with exploratory factor analysis. Question 4 was examined with item analysis for each of the subscales. The provision of a template adaptable instrument and a final step-by-step guide for adapting the instrument for other investigators answered question 5.

Cognitive Testing

The first research question was, "Will a thorough content validation process satisfy the needs for instrument validity?" The initial draft of questionnaire included three separate sections. Section one was the instruction for completing the questionnaire. Section two included demographic questions such as age, affiliation, specialty, etc. Section three included 32 items addressing the TPB constructs. The draft instrument can be found in Appendix D.

Several strategies were used to refine the instrument content in order to improve content validity. Cognitive testing was conducted with four physicians from the target population. The item modification suggestions were then collected and sent to the experts for review. The instrument was then revised based on experts' decisions in response to the results of the cognitive testing. After that, four other clinicians from the target population were selected to pilot test the instrument and then gave their comments. All the comments were then sent to the experts for review. The instrument was finalized with the experts' suggestions in response to the pilot testing results. The results of each of these steps are summarized below.

Four physicians from the target population were referred by the meeting organizer to conduct the cognitive testing for the instrument. Cognitive Testing

Instruction can be found in Appendix E. The clinicians were asked to walk through the instrument. They were asked to read the questions aloud to themselves and talk out loud about their reactions: if any of the questions were difficult to understand, were hard to answer, or did not make sense. They were also asked to answer the questions one by one and tell what were the responses they selected and why they selected them. Clinicians' suggestions to the draft instrument and the revisions made accordingly were listed in the following Table 4-1.

Based on all the suggestions for instrument questions and the responses from meeting organizers and experts for those suggestions, several revisions were made accordingly for the instrument. Both the initial draft instrument and the final instrument can be found in Appendix D and G.

The item "Your date of birth (mm/dd/yy)" was revised to "Your age" to assure privacy. The item "please indicate number of years in practice" was revised to "Please indicate number of years in Patient care ONLY; Research (non-patient care) ONLY; Patient care AND Research at the same time" to assure the accuracy of the answers. The decision was made to hand write numbers on the questionnaire. These numbers served as the questionnaire linking strategy to avoid confusion from trying to understand individuals' handwriting.

Table 4-1 Cognitive Testing Results

Section	Suggestions	Response
Age	Use "age" instead of "date of birth" to	Revised item "Your date of birth

	protect privacy.	(mm/dd/yy)" to "Your age".
Years of Practice in Research and Patient Care	The item "please indicate number of years in practice" was confusing. Instead, asking the number of years in patient care only, research only and both separately was suggested.	Revised the item to "Please indicate number of years in Patient care ONLY; Research (non-patient care) ONLY; Patient care AND Research at the same time".
Identifier for Linking	Using initials as linking method was not a good idea given the fact that it is often hard to understand individuals' handwriting.	Decided to hand write identification numbers on the instruments as the linking strategy
Item 13	Add "and deciding to participate in" after "developing" as not everybody was in the situation of developing research studies.	Changed "developing" to "developing and deciding to participate in" for item 13.
Item 15	"Credible—not credible" was not a good item.	Item was kept for the final draft to increase scale reliability.
Item 21	"Hard--easy" was not a good pair of words to use for item.	Item was modified to "complex--simple" as the introduced therapy complicates surgery.
Item 28	A time frame should be provided for the item.	Item was revised to "I intended to review literature about pre-operative systemic chemotherapy in the next two months".
Subjective Norm Scale	• Change "colleagues think I should" to "experts in the field think I should." • Add "experts" as an option to question "please indicate who influences your clinical decisions most." • Change "decision" to "decision-making" for question "please indicate who influence your clinical decision most."	• Changed "colleagues think I should" to "most clinicians whose opinion I value think I should…" • "Experts" was added as an option to question "please indicate who influenced your clinical decisions most". • "Decision" has been changed to "decision-making" for question "please indicate who influence your clinical decision most".
Overall	Delete "operable" from all items.	The word "operable" was kept for all the items in the instrument per experts' requests.
Format	The layout for the semantic scale's adjective words and the numbers between was confusing.	Adjective words were moved to above the seven numeric options instead of being aside (Appendix G).

Item 13 was revised from "developing" to "developing and deciding to participate in". This revision made the item applicable to more participants. Item 15 was kept for the final draft to increase scale reliability. Item 21 was modified from "hard--easy" to

"complex--simple" considering the introduced therapy complicates surgery. "Experts" was added as an option to question "please indicate who influence your clinical decision most." In the subjective norm scale, "Colleagues think I should" was changed to "most clinicians whose opinion I value think I should" for subjective norm scale since "experts in the field" instead of colleagues were considered to be the significant others for the clinicians in their clinical decision making. Likewise, "Decision" was changed to "decision-making" for the question "please indicate who influence your clinical decision most." A time frame of "in the next two months" was added to item 28 "I intended to review literature about pre-operative systemic chemotherapy" to facilitate the comparison. However, the word "operable" was kept for all the items in the instrument per experts' requests. Finally, adjective words for semantic scales were moved to above the seven numeric options to avoid confusion (Appendix F).

Pilot Test

CME evaluation instruments drafts with revised items from cognitive testing were administered to four clinicians from the target population. Since all the participants were not local, facsimile was used to collect the completed instruments. In addition to the instrument, participants also provided feedback to identify ambiguities and difficult questions. Time taken to complete the instrument was recorded and reported.

All the subjects reported using five minutes or less to fill out the questionnaire, suggesting the length of the questionnaire was reasonable. All the items in the instrument were unanswered as expected. No items were considered confusing, and the responses for the questions varied a lot, indicating that the questions were capable of differentiating among respondents. Thus, no further revisions were made as the result

of the pilot testing. The final draft instrument was sent to the meeting organizer and experts for final review before being administered during the conference.

This thorough content validation process including cognitive testing, experts' review and pilot testing satisfied the needs for a process to assure instrument validity. The instrument was shown to have a good content validity as each iteration required fewer adjustments from the comments of the target audience and expert reviewers.

Sample Size and Response Rate

According to the meeting organizer, the 155 CME evaluation forms were collected from 114 physicians and 41 non-physicians. There were 134 physicians and 30 non-physicians in our sample. Our sample was representative of the conference participants. In addition, there were 269 on-site participants so the response rate was 60.96%. Nurally (1978) suggested the need for 10 participants per item while Gorsuch (1983) suggested the ratio of 5 to 1. There were 22 items in our final scale and 134 valid participants in our sample. Therefore, the participant: item ratio was 6.09:1. This ratio meets Gorsuch's criteria.

Descriptive Statistics

Descriptive statistics for the demographic information (i.e. gender, specialty, affiliation, seeking CME credit) are provided in Table 4-2a and Table 4-2b. Descriptive statistics for the scale items and related factors (e.g. frequency, mean score, standard deviation, sample size) are provided in Table 4-3.

Table 4-2a. Demographic of the Sample Participating Physicians I

	Frequency	Valid Percent
Affiliation		
Academia	79.0	59.4
Government	19.0	14.3
Industry	3.0	2.3

Community Practice	27.0	20.3
Other	2.0	1.5
Academia and Community Practice	2.0	1.5
Academia and Government	1.0	.8
Total	133.0	100.0
Specialty		
Gynecologist	1.0	.7
Medical Oncologist	70.0	51.9
Surgeon	36.0	26.7
Pathologist	4.0	3.0
Radiation Oncologist	17.0	12.6
Radiologist	4.0	3.0
Surgical Oncologist	2.0	1.5
Endocrinologist	1.0	.7
Total	135.0	100.0
Seeking CME Credits		
Do not seek CME credits	26.0	19.7
Seek CME credits	106.0	80.3
Total	132.0	100.0
Gender		
Male	62.0	47.0
Female	70.0	53.0
Total	132.0	100.0

According to Table 4-2a, most of the sample physicians were affiliated with academic institutes (N=79, 59.4%), community practice (N=27, 20.3%) or government institutions (N=19, 14.3%), with some physicians having more than one affiliation. The total percentage of physicians affiliated with these top three categories was 96.3%. In terms of practicing specialty, most of the sample physicians were medical oncologists (N=70, 51.9%), surgeons (N=36, 26.7%), and radiation oncologists (N=17, 12.6%). This composition of participants was consistent with the target audience of the conference on preoperative breast cancer therapy. The majority of the sample of participating physicians were seeking CME credits (N=106, 80.3%). Male participants (N=62) represented 47% of the sample while female participants represented (N=70) 53% of the sample.

Table 4-2b. Demographic of the Sample Participating Physicians II

	Mean	SD
Age	47.7	8.93
Number of years in patient care only	3.8	7.73
Number of years in research (non-patient care) only	1.0	2.96
Number of years in patient care and research at the same time	12.2	10.13

According to Table 4-2b, the age of the sample physicians ranged from 26 to 69, with a mean age of 47.7. The number of years that physicians conducted patient care and research at the same time (Mean=12.2) was much higher than that of taking care of patients only (Mean=3.8) and that of conducting research only (Mean=1.0).

Means and standard deviations were calculated for each of the items in the instrument. The results revealed that most scale scores had mid to high ranges, indicating moderate to high levels of the underlying beliefs, attitudes, intentions, subjective norms and perceived behavioral controls. The constructs were represented in the theories' hypothesized directions.

Table 4-3. Descriptives for Items in the Instrument

									Mean	SD
Belief Items										
Pre-operative (as opposed to post-operative) systemic chemotherapy will										
improve breast conservation rates of operable breast cancer patients.									6.0	1.24
increase local recurrence rates of operable breast cancer patients.									4.9	1.71
increase disease-free survival rates for operable breast cancer patients.									2.8	1.62
increase the risk of inadequate surgery for operable breast cancer patients.									4.8	1.69
lead to a lower mortality rate in operable breast cancer patients.									2.5	1.49
reduce the overall medical costs for operable breast cancer patients.									2.6	1.37
have fewer side effects for operable breast cancer patients.									2.7	1.37
Attitude Items										
The practice of pre-operative (as opposed to post-operative) systemic chemotherapy is…										
Not credible	1	2	3	4	5	6	7	Credible	6.0	1.10
Unsafe	1	2	3	4	5	6	7	Safe	6.0	1.33
Harmful	1	2	3	4	5	6	7	Beneficial	5.3	1.53
Ineffective	1	2	3	4	5	6	7	Effective	5.3	1.49
Frustrating	1	2	3	4	5	6	7	Satisfying	5.2	1.46
Impractical	1	2	3	4	5	6	7	Useful	5.4	1.56
Complex	1	2	3	4	5	6	7	Simple	4.0	1.78
Perceived Behavioral Control Items										
Please rate your confidence level in…										
Sharing information about pre-operative systemic chemotherapy with operable breast cancer patients.									5.9	1.31
Sharing knowledge of pre-operative systemic chemotherapy with physicians who do not attend the conference.									5.9	1.21
Evaluating/assessing operable breast cancer patients' suitability for receiving pre-operative systemic chemotherapy.									5.7	1.30
Recommending pre-operative systemic chemotherapy to operable breast cancer patients.									5.6	1.46
Referring operable breast cancer patients to pre-operative systemic chemotherapy trials.									5.8	1.31
Applying knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher.									5.6	1.47
Evaluating pre-operative systemic chemotherapy papers critically when they appear in the literature.									5.6	1.31
Subjective Norm Items										

Most clinicians whose opinion I value think I should…		
Share information about pre-operative systemic chemotherapy with operable breast cancer patients.	5.6	1.36
Share knowledge of pre-operative systemic chemotherapy with physicians who do not attend the conference.	5.7	1.24
Recommend pre-operative systemic chemotherapy to operable breast cancer patients.	5.1	1.56
Refer operable breast cancer patients to pre-operative systemic chemotherapy trials.	5.5	1.32
Intention Items		
I Intend to…		
Share information about pre-operative systemic chemotherapy with operable breast cancer patients.	5.8	1.40
Share knowledge of pre-operative systemic chemotherapy with physicians who do not attend the conference.	6.1	1.03
Review literature about pre-operative systemic chemotherapy in the next month.	5.6	1.48
Apply knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher.	5.9	1.19
Evaluate pre-operative systemic chemotherapy papers critically when they appear in the literature.	6.2	1.05
Refer operable breast cancer patients to appropriate pre-operative systemic hormonal therapy trials.	5.2	1.67
Recommend appropriate pre-operative systemic hormonal therapy to operable breast cancer patients.	5.0	1.76
Scale Items		
Behavioral Intention	16.2	4.19
Attitudes	33.1	6.97
Subjective Norms	22.0	5.00
Perceived Behavioral Control	40.2	8.23

Construct Validity

Items (item 1-32) listed in the CME evaluation questionnaire that measured clinicians' behavioral intentions and their determinants were subjected to an exploratory factor analysis to determine the construct validity based on the Theory of

Planned Behavior. That information was used to answer the next two research questions: 2) Can an instrument for this purpose be developed with subscales in each of the theoretical domains; 3) Will a psychometric examination of the draft instrument reveal any unexpected measurement subscales?

Sampling Adequacy

Sample adequacy was tested by the Kaiser-Meyer-Olkin (KMO) statistic. Bartlett's Test of Sphericity test was also conducted to check the identity feature of the correlation matrix. Principal component analysis (PCA) was conducted to perform an initial extraction and to determine the number of factors being retained on the basis of multiple criteria, i.e. Kaiser's rule (extract all components with eigenvalues greater than or equal to one), Cattell's scree plot (look for big breaks in the scree plot), and Cattell -Nelson-Gorsuch's "objective" scree to seek the convergence among multiple criteria. Principal axis factor analysis (PAF) was then performed to extract the determined number of factors with oblimin rotation. Missing values were excluded using listwise deletion. Two criteria were used to determine whether an item was retained on a factor: 1) the factor loading was greater or equal to .50; and 2) if a variable loaded on more than one factor, the factor was retained on the factor with better conceptual consistency.

Table 4-4, KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.83
Bartlett's Test of Sphericity	Approx. Chi-Square	2871.12
	df	496
	Sig.	.00

Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy were used to evaluate the strength of the linear association among the 32 items in the correlation matrix. Bartlett's test of sphericity was significant (χ^2 = 2871.12, p=.00), which indicated that the correlation matrix was not an identity matrix. The Kasier-Meyer-Olkin (KMO) statistic was used to compare the magnitude of the observed correlation coefficients with the magnitude of partial correlation coefficients. The KMO coefficient (.83) was "meritorious" according to Kaiser's criteria (1960).

Further Individual Measure of Sampling Adequacy was used to determine the scale's factorability. With the exception of the second, fourth, and last correlation on the diagonal of the anti-image matrix, all other correlations ranged from 0.56 to .95. According to Kaiser's criterion, they are from "middling" to "marvelous." Overall, the results of the Bartlett's test, KMO, and MSA indicated that the correlations among the items warranted the PCA procedure.

Number of Factors Retained

Initial extraction was undertaken using unrotated PCA retaining as many factors as variables. Results of this initial extraction are presented in Table 4-5. Applying *Kaiser's Criterion* (Kaiser, 1960), which states that factors with eigenvalues greater than or equal to one should be retained, the first nine factors should be retained.

In addition, the Cattell-Nelson-Gorsuch "Objective" Scree test was also used to further determine how many factors needed to be retained. The results presented in Table 4-6 indicated that when the sixth component was added, there was a proportional decrease in the score, which also supported retaining six factors. Given that the results of the three procedures for determining the number of factors converged, six factors were retained for our factor analysis.

Table 4-5. PCA for Factors with Eigenvalues Greater Than 1

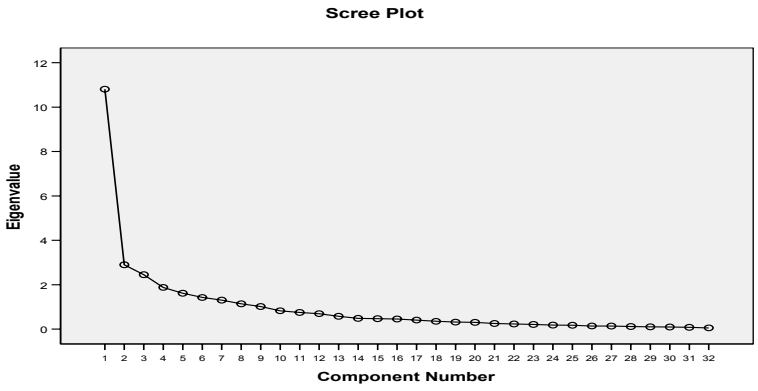
Factor	Eigenvalues	Percent of Variance	Cumulative Percent
1	10.81	33.78	33.78
2	2.90	9.06	42.84
3	2.45	7.66	50.50
4	1.88	5.87	56.38
5	1.62	5.06	61.43
6	1.43	4.46	65.89
7	1.31	4.08	69.97
8	1.14	3.57	73.54
9	1.02	3.18	76.72

Table 4-6. Cattell-Nelson-Gorsuch "Objective" Scree

Comparison	Scores
Comparing 1, 2 and 3	4.18
Comparing 2, 3 and 4	0.51
Comparing 3, 4 and 5	0.42
Comparing 4, 5 and 6	0.23
Comparing 5, 6 and 7	0.16
Comparing 6, 7 and 8	0.14
Comparing 7, 8 and 9	0.14

The results in the scree plot in figure 4-1 suggested that six factors should be retained. The plot shows there were six points beyond the line.

Figure 4-1. Scree Plot of the Principal Component Analysis



Comparing Initial and Extracted Communalities in PCA and PAF

Initial and extracted communalities in PCA and PAF among the six retained factors were compared (see Table 4-7). PAF results indicate that initial communalities in PAF were much less than 1 at range from 0.16 to 0.85.

Table 4-7 Communalities by Extraction Method (PCA & PAF)

	Principal Components Analysis		Principal Axis Factoring	
	Initial	Extraction	Initial	Extraction
CONSERVA	1.00	.79	.39	.16
RECURREN	1.00	.84	.57	.27
SURVIVAL	1.00	.73	.52	.32
INADSURG	1.00	.80	.61	.55
MORTALIT	1.00	.69	.60	.51
MEDCOST	1.00	.74	.59	.53
SIDEEFFE	1.00	.72	.63	.66
SESHINPA	1.00	.82	.86	.78
SESHKNPH	1.00	.82	.85	.75
SEEVSUIT	1.00	.84	.83	.85
SERECOMM	1.00	.80	.82	.75
SEREFER	1.00	.71	.71	.66
SEAPPLY	1.00	.68	.76	.59
SEEVALIT	1.00	.70	.69	.53
CREDIBLE	1.00	.71	.70	.56
SAFE	1.00	.75	.77	.68
BENE	1.00	.79	.80	.72
EFFECTIV	1.00	.77	.76	.70
SATISFY	1.00	.79	.77	.64
USEFUL	1.00	.74	.77	.66
SIMPLE	1.00	.72	.31	.13
SNSHINPA	1.00	.80	.88	.79
SNSHKNPH	1.00	.81	.88	.79
SNRECOMM	1.00	.80	.82	.71
SNREFER	1.00	.84	.79	.69
INTSHPA	1.00	.72	.77	.65
INTSHPH	1.00	.70	.71	.59
INTREVIE	1.00	.76	.54	.26
INTAPPLY	1.00	.69	.72	.56
INTEVALIT	1.00	.77	.69	.58
INTREFHO	1.00	.87	.79	.84
INTRECHO	1.00	.84	.73	.59

Extraction Method: Principal Component Analysis.

Table 4-8a and Table 4-8b present the total variance explained by six factors using PCA and PAF respectively. The six PCA factors accounted for 65.89% of the total variance whereas the three PAF factors accounted for 59.37% of the total variance. The fact that PAF extraction accounts for less variance than the PCA extraction illustrates an essential difference between the two extraction methods: PAF does not include unique variances while PCA does. Given that the initial communalities are much less than one, it was concluded that PAF is a better approach than PCA.

Table 4-8a Total Variance Explained (PCA)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Percent of Variance	Cumulative Percent	Total	Percent of Variance	Cumulative Percent
1	10.81	33.78	33.78	10.81	33.78	33.78
2	2.90	9.06	42.84	2.90	9.06	42.84
3	2.45	7.66	50.50	2.45	7.66	50.50
4	1.88	5.87	56.38	1.88	5.87	56.38
5	1.62	5.06	61.43	1.62	5.06	61.43
6	1.43	4.46	65.89	1.43	4.46	65.89

Extraction Method: Principal Component Analysis.

Table 4-8b Total Variance Explained (PAF)

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Percent of Variance	Cumulative Percent	Total	Percent of Variance	Cumulative Percent
1	10.98	34.18	34.18	10.61	33.16	33.16
2	3.00	9.36	43.54	2.60	8.13	41.29
3	2.44	7.61	51.15	2.12	6.63	47.92
4	1.91	5.98	57.13	1.38	4.31	52.23
5	1.62	5.05	62.18	1.20	3.76	55.99
6	1.44	4.49	66.67	1.08	3.38	59.37

Extraction Method: Principal Axis Factoring.
a When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Factors Generated Using Direct Oblimin Rotation

The Direct Oblimin rotation was applied to examine factor correlations. Table 4-9 Factor Correlation Matrix shows that all six factors were correlated with each other. The absolute values of the factor correlations ranged from .02 to .47, indicating that Direct Oblimin rotation should be used (Pett, Lackey, & Sullivan, 2003, p.165).

Table 4-9. Factor Correlation Matrix

Factor	1	2	3	4	5	6
1	1.00	-.17	-.36	-.02	-.28	-.47
2	-.17	1.00	.07	-.03	-.05	-.06
3	-.36	.074	1.00	.03	.15	.42
4	-.02	-.03	.03	1.00	-.07	-.05
5	-.28	-.05	.15	-.07	1.00	.38
6	-.47	-.06	.42	-.05	.38	1.00

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

Table 4-10 shows the factor loading matrix with PAF. We can see from the table that 25 items clustered onto six factors (values in boldface). Seven items (variable name and values in italics) did not load on any of the factors (subscales), suggesting that they should be eliminated from the instrument.

According to table 4-10, the first factor consisted of seven perceived behavioral control variables. The second factor consisted of three positive belief items. Five attitude items loaded on the third factor while two negative belief items loaded on the fourth factor. The fifth factor consisted of three intention variables.

Table 4-10 Factor Loadings with PAF

	Factor					
Variable	**Perceived Behavioral Control 1**	**Positive Beliefs 2**	**Attitudes 3**	**Negative Beliefs 4**	**Behavioral Intention 5**	**Subjective Norms 6**
CONSERVA	*.09*	*.12*	*-.08*	*-.09*	*-.01*	*-.27*
RECURREN	-.03	.11	-.13	**.50**	.12	.02
SURVIVAL	*-.17*	*.45*	*.03*	*.25*	*.06*	*-.14*
INADSURG	.15	.10	-.11	**.73**	.03	.06
MORTALIT	-.24	**.60**	.13	.08	-.07	-.06
MEDCOST	-.00	**.70**	.00	-.03	-.17	.18
SIDEEFFE	.19	**.81**	.11	-.01	-.02	-.03
SESHINPA	**.85**	.05	-.03	-.12	-.04	-.02
SESHKNPH	**.87**	.04	-.00	-.01	.08	-.05
SEEVSUIT	**.97**	.00	.07	.03	.08	-.00
SERECOMM	**.71**	.07	-.09	-.06	-.17	-.11
SEREFER	**.68**	-.07	-.14	-.02	-.03	-.08
SEAPPLY	**.59**	-.00	-.13	.03	-.15	-.10
SEEVALIT	**.70**	-.13	-.01	.14	.03	.03
CREDIBLE	*.21*	*.09*	*-.38*	*-.07*	*-.14*	*-.28*
SAFE	.20	-.03	**-.63**	-.08	-.14	-.09
BENE	.01	-.13	**-.79**	.18	.04	-.07
EFFECTIV	.03	-.09	**-.85**	.12	-.02	.13
SATISFY	-.04	.04	**-.69**	.00	-.07	-.21
USEFUL	-.02	-.07	**-.69**	-.06	-.18	-.13
SIMPLE	*.12*	*.08*	*-.12*	*-.21*	*.13*	*-.16*
SNSHINPA	.13	-.01	.03	.07	-.07	**-.80**
SNSHKNPH	.10	-.07	.03	.04	-.11	**-.80**
SNRECOMM	-.01	.00	-.17	.06	-.01	**-.75**
SNREFER	-.09	-.02	-.23	-.09	.04	**-.76**
INTSHPA	.26	.01	.08	.01	-.25	**-.54**
INTSHPH	*.29*	*-.11*	*.01*	*.19*	*-.23*	*-.42*
INTREVIE	*-.01*	*-.10*	*.13*	*.40*	*-.13*	*-.22*
INTAPPLY	.09	-.04	-.13	.17	**-.60**	-.09
INTEVALIT	*.42*	*-.15*	*.02*	*.25*	*-.37*	*-.10*
INTREFHO	-.09	.12	-.10	-.22	**-.86**	-.08
INTRECHO	.04	.18	-.09	-.05	**-.71**	-.00

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
a Rotation converged in 11 iterations.

The sixth factor consisted of four subjective norm variables and one intention variable. Although the item "intention to share information of preoperative breast cancer

therapy" loaded onto the subjective norm scale, it was not conceptually consistent with this scale. Therefore this item was not retained in the instrument to secure the measurement accuracy of the subscales. All the factors, factor (subscale) names, and the items loading on each factor are displayed in table 4-11 below.

Table 4-11. Factors Loadings for Items in Subscales

Factor	Subscale	Loadings	Questionnaire Item
1	Please rate your confidence level in…		
	Perceived Behavioral Control	.85	Sharing information about pre-operative systemic chemotherapy with operable breast cancer patients.
		.87	Sharing knowledge of pre-operative systemic chemotherapy with physicians who do not attend the conference.
		.97	Evaluating/assessing operable breast cancer patients' suitability for receiving pre-operative systemic chemotherapy.
		.71	Recommending pre-operative systemic chemotherapy to operable breast cancer patients.
		.68	Referring operable breast cancer patients to pre-operative systemic chemotherapy trials.
		.59	Applying knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher.
		.70	Evaluating pre-operative systemic chemotherapy papers critically when they appear in the literature.
2	Pre-operative (as opposed to post-operative) systemic chemotherapy will		
	Positive Beliefs	*.60*	*lead to a lower mortality rate in operable breast cancer patients. (not included in the final instrument)*
		.70	reduce the overall medical costs for operable breast cancer patients.
		.81	have fewer side effects for operable breast cancer patients.
3	The practice of pre-operative (as opposed to post-operative) systemic chemotherapy is…		
	Attitudes	-.63	Unsafe / Safe
		-.77	Harmful / Beneficial
		-.85	Ineffective / Effective

		-.69	Frustrating / Satisfying
		-.69	Impractical / Useful
4	Pre-operative (as opposed to post-operative) systemic chemotherapy will		
	Negative Beliefs	.50	increase local recurrence rates of operable breast cancer patients.
		.73	increase the risk of inadequate surgery for operable breast cancer patients.
5	I Intend to…		
	Behavioral Intention	*-.60*	*Apply knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher. (not included in the final instrument)*
		-.86	Refer operable breast cancer patients to appropriate pre-operative systemic hormonal therapy trials.
		-.71	Recommend appropriate pre-operative systemic hormonal therapy to operable breast cancer patients.
6	Most clinicians whose opinion I value think I should…		
	Subjective Norms	-.80	Share information about pre-operative systemic chemotherapy with operable breast cancer patients.
		-.80	Share knowledge of pre-operative systemic chemotherapy with physicians who do not attend the conference.
		-.75	Recommend pre-operative systemic chemotherapy to operable breast cancer patients.
		-.76	Refer operable breast cancer patients to pre-operative systemic chemotherapy trials.

Based on all the factor analyses results, we can see that a CME evaluation instrument was developed with each of the subscales representing a predetermined theoretical domain according to the Theory of Planned Behavior—i.e. attitudes, beliefs (including both negative and positive beliefs), perceived behavioral control (self-efficacy), subjective norms, and behavioral intention. In addition, a psychometric examination of this draft CME instrument by principal axis factoring with oblimin rotation revealed that the belief

items in the instruments belonged to two different subscales instead of one integrated belief scale. The two belief scales represented positive beliefs and negative beliefs, respectively.

Instrument Reliability

Item analyses were conducted for each of the revealed subscales in order to include the appropriate items. Item to total correlations and inter-item correlations were used to conduct the item analysis in order to decide which items should be retained for the final subscales. The developed multi-item scales are summated scales for attitudes, perceived behavioral control, behavioral beliefs (negative, positive), subjective norms, and behavioral intentions. Cronbach's alphas for each of the subscales were also examined to assess internal consistency reliability. The above information answers the following research question: Can an instrument be developed with acceptable levels of reliability for each of the necessary subscales?

Based on the results from the item analysis for all the six subscales revealed by factor analysis, a 22-item instrument is finally developed. Table 4-12 below shows the names of the subscales, their reliabilities and the number of items included in each subscale.

Table 4-12 Alpha Coefficients for Subscales in the CME Instrument

Factor	# Items in Scale	Factor Name	Standardized Item Alpha
1	7	Perceived Behavioral Control	0.94
2	2	Positive Beliefs	0.76
3	5	Attitudes	0.90
4	2	Negative Beliefs	0.74
5	2	Behavioral Intention	0.88
6	4	Subjective Norms	0.91

Table 4-13 Item-Total Statistics for Perceived Behavioral Control Scale

Variable	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Sharing information	34.09	47.58	**.83**	.84	**.92**
Sharing knowledge	34.08	48.89	**.83**	.84	**.93**
Evaluating suitability	34.27	47.25	**.86**	.79	**.92**
Recommending therapy	34.37	45.94	**.81**	.69	**.93**
Referring patients	34.12	48.26	**.79**	.66	**.93**
Applying knowledge	34.37	47.17	**.74**	.62	**.93**
Evaluating literature	34.29	49.45	**.71**	.60	**.93**

According to the above item-total statistics for perceive behavioral control subscale, a judgment was made about which items to retain for the final scale. The developed multi-item perceive behavioral control scale is a summated scale for the items in the scale. As shown in table 4-13, corrected item-total correlations ranged from 0.71 to 0.86. These were all high enough for the items to be retained for the subscale.

Inter-item correlations matrix can be obtained from computing the correlations for each pair of items. Inter-item correlations ranged from 0.54 to 0.88. As all the correlations in the matrix were higher than 0.5, it was concluded that all the items in this scale correlated well and should be retained in this subscale. The result was consistent with the item-total statistics.

A Cronbach's alpha of 0.93 of this subscale was obtained and considered to be a good value. In addition, all the alpha-if-item-deleted values in table 4-13 were lower than

or equal to 0.93, which confirmed the conclusion that no item should be eliminated from this subscale.

Table 4-14 Inter-Item Correlation Matrix Perceived Behavioral Control Scale

	Sharing info	Sharing knowledge	Evaluate suitable	Recommend	Refer patients	Apply knowledge	Evaluate literature
Sharing information	1.00						
Sharing knowledge	.88	1.00					
Evaluating suitability	.82	.83	1.00				
Recommending	.73	.68	.74	1.00			
Referring patients	.71	.65	.72	.73	1.00		
Applying knowledge	.59	.61	.61	.70	.66	1.00	
Evaluating literature	.54	.61	.68	.60	.61	.69	1.00

The summated scale measured the perceived behavioral control for performing the pre-operative breast cancer therapy. The higher the score value, the stronger perceived behavioral control clinicians had for adopting the clinical practice introduced in the CME conference (pre-operative breast cancer therapy), which in turn, is supposed to predict a higher intention of performing this clinical practice.

Table 4-15 Item-Total Statistics for Positive Belief Subscale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Decreased mortality	5.25	6.07	**.46**	.22	**.76**
Lower medical cost	5.10	5.78	**.61**	.41	**.58**
Fewer side effects	5.04	5.83	**.60**	.40	**.59**

According to the above item-total statistics for the positive belief subscale (see Table 4-15), corrected item-total correlations ranged from 0.46 to 0.60. The item "Pre-operative (as opposed to post-operative) systemic chemotherapy will lead to a lower mortality rate in operable breast cancer patients" had the lowest corrected item-total correlation of 0.46. This was much lower than that of other items. While in an acceptable range, this item was not as closely associated with the rest of the scale as the other items.

Table 4-16 Inter-Item Correlation Matrix Perceived Positive Belief Subscale

	Decreased mortality	Lower medical cost	Fewer side effects
Decreased mortality	1.00		
Lower medical cost	.42	1.00	.61
Fewer side effects	.41	.61	1.00

Inter-item correlations for positive beliefs ranged from 0.41 to 0.61. All the correlations in the matrix are higher than 0.2, so correlations existed among all the items in this subscale. Although two of them were lower than 0.5 needed to be considered good correlations, they were still acceptable while higher than the criteria of divergent correlation. Consideration may be given to either remove or revise the item. The result was consistent with the item-total statistics.

A Cronbach's alpha of 0.73 of this subscale was considered to be a good value. One of the alpha-if-item deleted values was higher than the scale alpha. The increased alpha value was 0.76, and the deleted item was "Pre-operative (as opposed to post-operative) systemic chemotherapy will lead to a lower mortality rate in operable

breast cancer patients." This result was consistent with the previous results from item-total correlations and inter-item correlation matrix. Given the consideration from all three results, the decision was made to drop this item from this subscale. The alpha value for the revised scale was 0.76.

The summated scale measured the positive beliefs for performing the pre-operative breast cancer therapy. The higher the score value, the more positive beliefs clinicians had for adopting the clinical practice introduced in the CME conference (pre-operative breast cancer therapy), which in turn, is supposed to predict a higher intention of performing this clinical practice.

Table 4-17 Item-Total Statistics for Attitude Subscale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
SAFE	21.15	27.47	**.70**	.52	**.89**
BENE	21.81	24.86	**.77**	.71	**.87**
EFFECTIV	21.74	25.51	**.74**	.66	**.88**
SATISFY	21.91	25.55	**.77**	.67	**.87**
USEFUL	21.69	24.77	**.76**	.68	**.87**

The above item-total statistics for attitude subscale were used to determine which items to retain in the final scale. The multi-item perceived behavioral control scale is a summated scale. As shown in Table 4-15, corrected item-total correlations ranged from 0.70 to 0.77, which are high enough for all the items to be retained in the subscale.

Table 4-18 Inter-Item Correlation Matrix for Attitudes Subscale

	Safe	Beneficial	Effective	Satisfying	Useful
Safe	1.00				
Beneficial	.62	1.00			
Effective	.54	.79	1.00		
Satisfying	.60	.65	.59	1.00	
Useful	.66	.56	.59	.78	1.00

The inter-item correlations from the above matrix ranged from 0.54 to 0.79. They were all above the 0.5 criterion for good correlation. This matrix suggested that all the items in this subscale correlated well and should be retained. The result was consistent with the item-total statistics.

A Cronbach's alpha of 0.90 of this subscale was obtained and considered to be a good value. In addition, all alpha-if-item-deleted values in Table 4-15 were lower than or equal to 0.93, which confirmed the conclusion that no item should be eliminated from this subscale.

The summated scale measured the attitudes toward performing the pre-operative breast cancer therapy. Higher scores indicate more favorable attitudes that clinicians have toward adopting the clinical practice introduced in the CME conference (pre-operative breast cancer therapy), which in turn, is supposed to predict a higher intention of performing this clinical practice.

Table 4-19 Item-Total Statistics for Negative Beliefs

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Recurrence	4.76	2.84	**.59**	.34	.(a)
Inadequate surgery	4.93	2.94	**.59**	.34	.(a)

According to the above item-total statistics for the negative beliefs subscale (see Table 4-19), corrected item-total correlations were 0.59, which was higher than the 0.5 criterion of good correlations. Since only two items were in the scale, 0.59 is also the

inter-item correlation. Cronbach's alpha for this two-item scale is 0.74, which was good enough and consistent with the inter-item correlation and item-total correlations.

The summated scale measured the negative beliefs for performing pre-operative breast cancer therapy. The higher the score value, the fewer negative beliefs clinicians have for adopting the clinical practice introduced in the CME conference (pre-operative breast cancer therapy), which in turn, is supposed to predict a higher intention of performing this clinical practice.

Table 4-20 Item-Total Statistics for Behavioral Intention

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Apply knowledge	10.12	10.51	**.50**	.26	**.88**
Refer trials	10.85	6.37	**.79**	.65	**.59**
Recommend therapy	11.02	6.19	**.74**	.63	**.64**

The item-total statistics for behavioral intention were used to determine which items to retain for the final scale (see Table 4-20). The corrected item-total correlations ranged from 0.50 to 0.79. Although they were all relatively high, the value for item "intention to apply knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher" (0.50) was much lower than those of the other two intention items in the subscale and should be considered either for revision or elimination from the scale. In addition, according to the inter-item correlation matrix listed below, the inter-item correlations between the above scale item with the other two (0.45 and 0.50) were much lower than the inter-item correlations between the other two—0.79, which also suggested this item should not be retained for this subscale. A

Cronbach's alpha- if-item-deleted of 0.88 was also higher than the original scale alpha of 0.81. All the above information are consistent and suggested the need to eliminate the item "intention to apply knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher" from the scale. The new Cronbach's alpha value became 0.88 after eliminating the item and rerunning the analysis for the new subscale.

Table 4-21 Inter-Item Correlation Matrix for Behavioral Intention Subscale

	Apply knowledge	Refer trials	Recommend therapy
Apply knowledge	1.00		
Refer trials	.50	1.00	
Recommend therapy	.45	.79	1.00

The summated scale measured the behavioral intention for performing the pre-operative breast cancer therapy. Higher score values indicate higher clinician intentions for adopting the clinical practice introduced in the CME conference (pre-operative breast cancer therapy).

The item-total statistics for subjective norms subscale in Table 4-22 shows all corrected item-total correlations were high, ranging from 0.76 to 0.83. No item was eliminated from this subscale.

All inter-items correlations shown in the above matrix were higher than the criterion of 0.5, indicating good correlations among the items in the subscales. In addition, the Cronbach's alpha for this scale was 0.91, higher than all the Cronbach's

alpha-if-item-deleted values listed in Table 4-22. All the results are consistent with each other and confirmed the conclusion that all the items should be retained in this subscale.

Table 4-22. Item-Total Statistics for Subjective Norms Subscale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Share information	16.28	13.53	**.83**	.83	**.87**
Share knowledge	16.16	14.49	**.81**	.81	**.88**
Recommend therapy	16.78	12.60	**.78**	.67	**.89**
Refer trials	16.39	14.33	**.76**	.63	**.89**

Table 4-23 Correlation Matrix for Subjective Norm Subscale

	Share information	Share knowledge	Recommend therapy	Refer trials
Share information	1.00			
Share knowledge	.90	1.00		
Recommend therapy	.69	.64	1.00	
Refer trials	.64	.64	.77	1.00

The summated scale measured the subjective norms for performing the pre-operative breast cancer therapy. Higher score values would indicate stronger subjective norms among clinicians for adopting the clinical practice introduced in the CME conference (pre-operative breast cancer therapy), which in turn, is supposed to predict a higher intention of performing this clinical practice.

Guidelines for Using Evaluation Instrument in Future CME Evaluation

Two separate sets of work have been done to answer the following question: Will a thorough instrument development process result in an instrument that is appropriate for evaluation of CME? The first set of work focused on whether a CME instrument could be developed to evaluate the adjuvant therapy for breast cancer conference. This was answered through the results of cognitive testing, expert opinion, pilot testing, and the data analysis for the conference presented in the aforementioned sections answering all the previous four research questions. The second part could be interpreted as whether the questions themselves with their corresponding constructs will be adaptable to evaluate other CME conference by revising the content.

An adaptable instrument template and a step-by-step guide for adapting the instrument for other CME conferences had been developed as follows. As the instrument has not been applied to other CME conferences, this part of the question can only be answered by future research.

Figure 4-2 Final Instrument Template

Demographics

Please circle your specialty

Please circle your dominant affiliation	Academia Community Practice	Government Industry Other (Please specify) _____
Please indicate number of years in	Patient care ONLY: _____ Research (non-patient care) ONLY: _____ Patient care AND Research at the same time: _____	
Your age:		
Do you seek CME credits?	YES	NO
Your gender:	FEMALE	MALE

Positive Behavioral Beliefs

[Medical procedure] will have *fewer side effects* for [type of] patients.	Unlikely (1)—likely (7)
[Medical procedure] will *reduce the overall medical costs* for [type of] patients.	

Negative Behavioral Beliefs

[Medical procedure] will *increase the risk of harmful medical procedure* for [type of] patients.	Unlikely (1)—likely (7)
[Medical procedure] will *increase the vicious clinical results* for [type of] patients.	

Attitudes

The [Medical procedure] is…								
Unsafe	1	2	3	4	5	6	7	Safe
Harmful	1	2	3	4	5	6	7	Beneficial
Ineffective	1	2	3	4	5	6	7	Effective
Frustrating	1	2	3	4	5	6	7	Satisfying
Impractical	1	2	3	4	5	6	7	Useful

Subjective Norms

Most clinicians whose opinion I value think I should…	Unlikely (1)—likely (7)
My colleagues think I should share information about [Medical procedure] with the [type of] patients.	
My colleagues think I should share knowledge of [Medical procedure] with physicians who do not attend the conference.	
My colleagues think I should recommend [Medical procedure] to the [type of] patients	
My colleagues think I should refer the [type of] patients to the trails of [Medical procedure]	

Behavioral Intention

I intend to	Unlikely (1)—likely (7)
provide [medical procedure] to [type of] patients in [specific instance].	
refer [type of] patients to appropriate trials in [medical procedure].	

Perceived Behavioral Control (Self-Efficacy)

Rate your confidence level in	
sharing information about [medical procedure] with [type of] patients.	Unconfident (1) --- Confident (7)
sharing knowledge of [medical procedure] with physicians who do not attend the conference.	
evaluating/assessing [type of] patients' eligibility of receiving [medical procedure] in [specific instance].	
recommending [medical procedure] to [type of] patients in [specific instance].	
providing [medical procedure] to [type of] patients in [specific instance].	
referring [type of] patients to appropriate trials in [medical procedure].	
applying knowledge of [medical procedure] in developing research studies.	
evaluating [medical procedure] papers critically when they appear in the literature.	

Instrument Adaptation Protocol

Figure 4-3 Instrument Development Protocol

1. Specialize the items in the template according to the CME conference objectives.
2. Conduct cognitive testing with small sample from target population.
3. Revise the daft survey based on the comments from the cognitive testing sample and the responses to the comments from conference experts.
4. Pilot test the revised instrument with small sample from target population.

An instrument development protocol was developed as above (Figure 4-3). CME activities vary greatly by clinical domains. Therefore, CME conference organizers

must first modify the items in the template to address the specific medical content focused by their CME efforts. Experts in the related medical fields should be consulted in order to develop the questions. Second, cognitive testing with a sample of nine or fewer members of target population is recommended to ensure the consistency between the information written in the instrument and the message delivered to the participants. Third, draft instruments should be revised based on the cognitive testing results and the expert feedback to those comments. Last, another nine or fewer clinicians from the target population should be asked to complete the revise the survey to identify ambiguities and difficult questions. At this pilot test, time taken to complete the instrument should also be recorded and reported. Future evaluation efforts that adapt the evaluation instrument template may yield information about whether and how this process can be streamlined.

Summary

This chapter reported the characteristics of the conference participants and the results of content validity, construct validity, and reliability analyses. Demographic information presented for the participants completing the instrument included academic affiliations, age, gender and clinical specialty. This information helped investigate the representativeness of the sample. The content validation processes included cognitive testing, expert review, and pilot testing. The construct validity and reliability analysis included factor analysis from the instrument, item analysis, and internal consistency analysis for all the subscales revealed by factor analysis. The results of analyses were used to answer the research questions.

Chapter 5: Discussion, Recommendations, and Conclusions

Study Summary

This chapter presents a summary of the purpose, methodologies, results, conclusions, discussion, and recommendations from the study. The current study was conducted due to the lack of availability of valid and reliable instruments to evaluate the effectiveness of CME activities. The evaluation of physician behavior change (level 3) and patient outcomes (level 4) should always built on the solid measures of attitude and belief change from level 2 (Kirkpatrick, 1994). The determination of the key factors for the success of CME activities was limited by the heterogeneity of CME programs in previous studies. The reliability and validity of the instruments that have been used to assess CME effectiveness limit the evidence (AHRQ, 2007). The lack of a theory base for the existing CME evaluation instruments addressing the variables in the second evaluation level has also limited the findings from previous CME evaluation studies (Tian et al, 2007).

This study determined the feasibility of adapting a theory-based instrument template to an NCI CME conference that would result in a content-specific, valid and reliable CME evaluation instrument assessing the changes in the concepts listed in the second evaluation level (i.e. attitudes, beliefs, subjective norms, perceived behavioral controls and behavioral intentions). The established and validated instrument provides evidence that adapting the theory-based template can be used in other evaluations of CME activities addressing physician clinical practice change.

Theoretical constructs that have been demonstrated to predict health and clinical practice behavior were integrated and applied (Ajzen, 1991; Bandura, 1982, 1989; Glanz, 1997). This involved applying Ajzen's Theory of Planned Behavior [TPB] (1991),

Bandura's Self-efficacy Theory (1982), and Bandura's Social Cognitive Theory [SCT] (1989). The application of perceived behavioral control in the TPB was consistent with Glanz et al's (1997) suggestion that the construct of perceived behavioral control was similar to the construct of self-efficacy by Bandura (1991). Therefore, perceived behavioral control served as a proxy measure of self-efficacy in this instrument. The behavioral beliefs construct in the TPB was used as a proxy measure for outcome expectations in the SCT in this instrument since these two constructs parallel each other.

Most of the constructs being evaluated in the instrument (belief, attitude, perceived behavioral control and subjective norms) belong to the second evaluation level of Kirkpatrick's Outcome Evaluation Model (1994). Behavioral intention, however, is located between the second and third evaluation levels as it is determined by the constructs from the second level while predicting the element (behavior) in the third evaluation level. Behavioral intention questions can serve as a proxy measure of physician behavior in the third evaluation level.

A thorough content validation process was used to examine the content validity of the instrument and answer the first research question "Will a thorough content validation process satisfy the needs for instrument validity?" The validation process included cognitive testing of the draft instrument with four clinicians from the target population, expert review and comments about cognitive testing results, pilot testing of the revised draft with another four clinicians from the target population, and finalizing the instrument based on the previous activities. The time needed to fill out the instrument was less than five minutes.

After obtaining IRB approval, consent forms along with traditional (paper and pencil) self-administered instrument were distributed in the conference registration package. Clinicians were encouraged to participate before the conference began both verbally and visually. A total of 164 questionnaires were returned, yielding a response rate of 60.96%. Non-physician cases (N=30) in the dataset were filtered out since this research targeted physicians only, resulting in 134 valid cases.. The sample characteristics suggested that the sample was representative of the target physician population.

Factor analysis with principal axis factoring with oblimin rotation was conducted to examine the construct validity of the instrument. Six instead of five factors were extracted from the 32 items of the instrument. Belief items loaded onto two different subscales: positive belief scale and negative belief scale. All the other subscales were in the predetermined theoretical domains, i.e. attitudes toward the behavior, subjective norms, perceived behavioral control, and behavioral intention. Seven of the thirty-two items did not load high enough on any of the factors and were eliminated from the instrument. Twenty-five items were retained in the instrument after the factor analysis.

Item analyses were conducted for each of the six revealed subscales in order to examine the internal consistency reliabilities of those subscales. Cronbach's alphas for all the subscales were also examined demonstrating adequate reliabilities ($0.73 < \text{alpha} < 0.94$). As a result of this process, three more items were dropped from the 25-item instrument, and the final established instrument had 22 items with six subscales.

Part of question 5 was answered by the results of cognitive testing, expert opinion, pilot test, and the data analysis in developing the instrument specialized for this NCI conference. Overall, the CME evaluation instrument for the NCI breast cancer

conference appears to have sufficient validity and acceptable levels of reliability for early instrument development. Given the expertise of the physicians' review and comment on the instrument, the instrument appeared to have strong face validity. The measures represented the general domains of the constructs. The sample physicians selected moderate to high response score for all the constructs in the instrument. The other part of the question, however, can only be answered by future research. This question is regarding whether the items themselves with their corresponding constructs can be adapted to evaluate other CME conferences addressing different clinical domains. An adaptable instrument template and a step-by-step adaptation guide have been developed to guide adaptation by other CME conferences for use in evaluating their intervention effectiveness.

Discussion of Results

Significant findings in the data analyses conducted to answer the research questions are the foci of the discussion of results. Although this instrument is unique for breast cancer clinicians receiving didactic CME intervention, modifications could be made to have broader implications. The results can be generalized to other CME activities with caution. According to a systematic review of CME evaluation studies by Tian and colleagues (2007), CME activities vary greatly by study design, intervention strategy, length of follow-up, clinical domain, and target audience. Special considerations are required when interpreting study results and in generalizing results to other CME activities.

Sample Characteristics

The analysis of the demographics of the sample physicians revealed that the sample was diverse in terms of academic affiliations and specialties while balanced on gender. The composition of the sample appeared to be representative of the physicians and non-Physicians who submitted NIH CME questionnaires to the conference organizer. The majority of the participants were medical oncologists, surgeons, or radiation oncologists. The sample composition indicated that most of the physicians attending the conference were medical oncologists affiliated with academic institutes. This reflected the purpose of this CME conference "Preoperative Therapy in Invasive Breast Cancer" and was representative of its target population. The sample had a relatively balanced gender composition.

A high response rate is the key to assure the accuracy of a survey's results because the sample is more likely to represent the overall target population. Low response rates, on the other hand, can damage the credibility of a survey's results. The current study had a response rate of 61%. According to Babbie et al. (2000), a response rate of 50% is adequate for analysis and reporting, a response rate of 60% is good, and a response rate of 70% is considered to be very good. The 61% response rate of this study is considered to be "good."

The guidelines for minimum ratios of participants to items proposed by Gorsuch (ratio=5:1, 1983) and by Nurally (ratio= 10:1, 1978) has been widely cited in psychometric analyses. There were 22 items in our final scale and 134 valid participants in our sample, resulting in a participant: item ratio of 6:1. Although this ratio was not very high, it was still acceptable and met Gorsuch's minimum ratio criterion. Both the

good response rate and the acceptable participant to item ratio supported the credibility of the study results.

Content Validity

The usage of expert review and cognitive testing methods to validate the questionnaire items was one of the advantages of this instrument development research. According to a recent review of 32 randomized CME evaluation studies, half used surveys/questionnaires without psychometric testing for validity and reliability, and only six studies adopted existing instruments in their field that had reliability and validity information (Tian et al, 2007). In addition, no cognitive testing methods were found in any of those instrument development processes (Sanci et al., 2000, Sanders et al., 2003, Jacobs et al., 2005)

This process supported the recommendation that expert review and pilot and cognitive testing be a standard part of any survey instrument development process (Collins, 2003). Expert review enabled the identification of the most acceptable response scale format. The use of cognitive testing methods in this research enabled us to explore the clinicians' question answering processes and the factors influencing their answers. Pre-testing questions in the questionnaire context and following modifications assured that future participating clinicians could understand the question concepts, in a consistent way, and in a way that was intended. In other words, cognitive testing methods assured that instrument items were measuring the theoretical concepts as was intended among participating physicians in a consistent manner. For example, the feedback obtained enabled the revision of subjective norm-related questions to include experts as important individuals that influence clinical practices.

Pilot testing showed that the instrument completion time was less than five minutes—very reasonable for the conference administration. Conference organizers were very concerned that completing the instrument could be done efficiently. Physicians participating in the pilot study also assured flow, salience, ease of administration, and acceptability of the revised instrument.

Construct Validity

The findings from the factor analysis demonstrated that interpretable factors existed in the CME evaluation instrument. In addition, only seven of the thirty-two items were not retained for use in the future reliability analysis. These findings suggested that the subscales revealed by the factor analysis were valid, which answered the research questions.

The suppressing criterion of factor loadings used in this study was 0.50. In other words, items would be considered retained in a subscale only if its loading on that factor was higher or equal to 0.50, which was more rigorous than typical loading criteria of 0.3. The loading of 0.5 indicates the item accounts for 25% of the scale variance. Despite such rigorous loading criteria, many of the *a priori* items were retained on the factors they were designed to address.

The *a priori* items in the initial instrument clustered together in a logical way for four of the six TPB constructs as expected: perceived behavioral control, attitudes, subjective norms, and behavioral intention.

One of the behavioral intention items loaded on the subjective norm scale, but.it was eliminated from the subjective norm scale in the final instrument due to the following reasons. First, the loading of 0.54 for this intention item is “fair” while the loadings of

the other four subjective norm items in the scale were all considered "excellent" (Comrey and Lee, 1992). Second, it was suggested that a definitive interpretation of the factor could be confidently achieved with several items' loadings on this factor being classified as "very good" (i.e. loadings > 0.63) or "excellent" (i.e. loadings > 0.71). The four "excellent" loadings from the subjective norm items in the scale provided sufficient evidence of its validity. Third, given the special feature of target population (clinicians) of CME evaluation instrument and the onsite administering strategy, eliminating this item enabled the development of a shorter instrument, which are preferred.

Belief items loaded separately on two factors, i.e. "positive beliefs" and "negative beliefs," which was not as expected. According to TPB, attitude is determined by the behavioral belief score (rated from -3 to +3) weighted by an evaluation of each individual belief (rated from -3 to +3). This arrangement captures the psychology of double negatives (Glanz, 1997). In the current study, however, the beliefs were numbered from 1 to 7, based on the expert review. No bi-polar component was considered in the scoring system for the current study, although all the responses to the negative belief items were reverse coded to be consistent with the responses to the positive beliefs directionally. In other words, a higher score on positive beliefs and a lower score on negative beliefs relates to a more positive attitude toward adapting the pre-operative breast cancer therapy.

Beliefs can be positive or negative about a new clinical practice. In general, TPB does not separate positive and negative beliefs into different constructs, and previous CME belief scales have had only one subscale. However, the decisional balance construct in the Transtheoretical model (Prochaska & DiClemente, 1983) provides a rationale for our results. This model differentiates between pros and cons of behavior

change and reflects the individual's relative weighing of them (Prochaska, et al. 1994). This model suggests that having separate positive belief and negative belief scales in this research was appropriate.

The other four subscales also had evidence for validity. The seven items in the perceived behavioral control scale had "good" to "excellent" loadings. Both of the intention items also had "excellent" loadings, and the attitudes scale had "excellent" and "very good" loadings. All subjective norm items had "excellent" loadings. According to Comrey and Lee's criteria (1992), these subscales could be condidently named as perceived behavioral control, attitudes, and subjective norms. These results also supported that the aforementioned three subscales were reliable (Guadagnoli & Velicer, 1988).

Only three items each loaded on the behavioral intention scale and the negative belief scale, and two items loaded on the positive belief scale initially. The sample size of this study (N=134) was not sufficient to support the interpretability of these three scales considering Guadagnoli and Velicer's suggestion (1988), which states that factors with about 10 or more loadings (around 0.4 in absolute value) should be considered reliable, as long as sample size is greater than about 150, and factors with only a few loadings should not be interpreted unless the sample size is at least 300. The initial instrument and subscales had a limited number of items available for factor analyses (item number=32, factor number=6), and four items for each subscale is the best situation that could be made except for the perceived behavioral control subscale. The decision to limit the number of items in the initial instrument/subscales was made because of the participants (i.e. physicians) and the on-site instrument administering strategy. Therefore, the use and

interpretation of the three 2-item subscales—positive beliefs, negative beliefs, and behavioral intention—was justified.

Reliability

Overall, all subscales had sufficient reliability (alpha>= 0.65) for early stage of instrument development (Nunnally et al., 1994) showing the unidimensionality of the subscales. In addition, scale modifications for this study were based on item analyses by considering item-total correlations, inter-item correlations and alpha-if-item-deleted values. For subscales with borderline Cronbach's alpha coefficients, problematic items were eliminated, and the analyses were rerun. Our results reflect these changes, while having a strong content validity maintained. Based on the results from the item analyses for all six subscales resulting from the factor analysis, a 22-item instrument was finally developed.

The perceived behavioral control subscale contained seven items with a Cronbach's alpha coefficient of 0.94. All the inter-item correlations indicated good correlations among the items within the scale (Trochim, 2001), so all seven items were retained in this subscale according to the item analysis results.

The positive belief scale contained three items with a Cronbach's alpha coefficient of 0.73. Two of the three inter-item correlations were not considered as acceptable (Trochim, 2001). Eliminating one of the three items [i.e. Pre-operative (as opposed to post-operative) systemic chemotherapy will lead to a lower mortality rate in operable breast cancer patients] from this subscale and rerunning the item analysis improved the alpha value to 0.76.

The attitude subscale contained five items with a Cronbach's alpha coefficient of 0.90. All the inter-item correlations met the criteria, so all five items were retained in this subscale.

The negative belief subscale contained two items with a Cronbach's alpha coefficient of 0.74. The two items had a good inter-item correlation of 0.54, so both items were retained in this subscale.

The behavioral intention subscale contained three items with a Cronbach's alpha coefficient of 0.81. Two of the three inter-item correlations did not meet the criteria. The subscale's improved alpha value improved to 0.88 after eliminating one of the three items (i.e. intention to apply knowledge of pre-operative systemic chemotherapy in developing or deciding to participate in research studies as a researcher).

The subjective norm subscale contained four items with a Cronbach's alpha coefficient of 0.91. All four items were retained in this subscale according to the inter-item correlations and item analysis results.

Cronbach's alpha is positively related to the number of items in the scale. Increasing the number of items in the scale will increase the reliability value even with the same degree of inter-correlation (Hair, 1998). However, two more items were eliminated from the subscales of positive beliefs and behavioral intention. Cronbach's alpha coefficients for both subscales increased after eliminating the problematic items, and inter-item correlations within both subscales increased after eliminating the problematic items. Eliminating items also allowed a shorter instrument, which is generally preferred by clinicians, according to the experts and CME conference organizers.

The items analyses conducted for the six subscales and the scale modifications based on the results were an advantage of this research. Although three pre-existing CME evaluation questionnaires that address the variables in the second evaluation level provide validity and reliability information, item analyses methods were not apparent in any of those instrument development processes to examine the scale reliability (Sanci et al., 2000, et al., 2003, Jacobs et al., 2005). Instead, only the Cronbach's alpha coefficients were provided as evidence for their reliabilities.

Limitations

The present study has limitations concerning sampling, participant to item ratio, scale characteristics, and data characteristics. Larger sample sizes are likely to result in more stable correlations among variables and in greater replicability of exploratory factor analysis outcomes. Given the good response rate (60.9%) of the study, small sample size is one reason for this relatively low participant to item ratio. One important reason for this small sample size was the use of a synchronized Internet broadcast of the conference. According to the meeting organizer, 431 participants registered for the conference, but only 269 actually attended the conference on site. This produced an attendance rate of 62.41%. If all the registrants attended the conference on site, the participant to item ratio could have increased to 10:1.

Participants not only had diverse academic affiliations and specialties, but they also came from several different countries. This information was obtained from one participant accidentally and was further confirmed by the conference organizer. Had the international feature been made clear in advance, a nationality or resident country question could have been included in the questionnaire. The nationality or the resident country of

the participants could potentially confound the study results because some European countries (e.g. Austria) do not require CME credits for physicians. The organizer's physician/non-physician composition data came from those seeking CME credits only. Although the sample's physician/non-physician composition was very similar to that of those seeking CME credits, one fifth of the participants in our sample did not seek CME credits. We do not know the participant composition of those who did not submit the CME evaluation forms to the meeting organizer and, therefore, could not compare them with our sample.

Pre-operative breast cancer therapy is a potentially controversial topic. During cognitive testing, one of the physicians who participated stated, "I am strongly against pre-operative breast cancer therapy, so that I am even not going to the conference." The controversial situation of this topic may lead to the selection bias of the participants. In other words, most of physicians who attended the conference were likely supportive of pre-operative breast cancer therapy or may have already been practicing preoperative therapy. Therefore, their responses to the items in the questionnaire cannot represent those provided by physicians who might be against this therapy. On the other hand, the conference might also potentially increase uncertainty about preoperative therapy among attendees by addressing the negative outcomes of this therapy, as showed in our negative belief scale. When generalizing the results from this research, selection bias of the sample needs to be considered. A similar situation might also exist within other clinical domains or topics. Instruments should be adapted to these domains/topics with caution.

As far as the scale characteristics, three constructs of the scale were represented by a small number of items. Cronbach's alpha coefficients factored into the equation the

number of items in the scales. Although eliminating the problematic items was justified in the aforementioned discussion for the current scale, modifications could be made for the problematic items in the future in order to increase the number of items per subscales. Given the short completion time for this questionnaire, adding two more items for each of these three subscales would not be unreasonable.

Only one administration method (paper-pencil) was applied in this research, so the Multitrait-Multimethod Matrix could not be developed to assess the convergent and discriminant validity for this instrument. The approach could provide more evidence for construct validity. In addition, there were no criterion validity analyses and theory testing analyses conducted for this instrument.

Another limitation was that the data were obtained through self-reported questionnaires. According to the literature, physicians usually overestimate their behavior (Davis et al., 2006), and social desirability bias may be evident in the data collected. For example, the clinicians might have indicated a higher behavior intention of practicing pre-operative breast cancer therapy than they truly would.

Although this instrument was adapted from the CME evaluation instrument template, it was designed uniquely for breast cancer clinicians receiving didactic CME intervention. Special considerations are required when interpreting the study results and generalizing the results to other CME activities.

Recommendations for Future Studies

The current study was able to develop a clinical-domain specific CME evaluation instrument adapted from an instrument template, examine the validity and reliability of this instrument, and revise the template from the validity and reliability results. This development process also serve as a model for creating other CME evaluation instruments. Several future lines of inquiry have been suggested by the findings from this study.

A larger sample size in future studies using the adapted instrument could increase the participant: item ratio, increase the interpretability of the factors with few loadings, and increase the feasibility of items analyses for the subscales, which in turn, would lead to a higher credibility of the study results. As discussed before, one important reason for the small sample size in the current study was the use of a synchronized Internet broadcast of the conference. Similar situations might also exist in future CME activities, which suggests the need to implementing data collection strategies other than on-site paper-pencil to capture those participants using the Internet conference broadcast. An online web-based gate-keeping survey could be an option.

Moreover, the online web-based gate-keeping survey for those physicians accessing the conference from the Internet could potentially help build a Multitrait-Multimethod Matrix. Multitrait-Multimethod Matrix is an approach to assessing the construct validity of a set of measures in a study (Trochim, 2001). Two subcategories of construct validity could be assessed with this matrix, namely; convergent validity—the degree to which concepts that should be related theoretically are actually interrelated—and discriminant validity—the degree to which concepts that should not be related theoretically are not interrelated in reality (Trochim, 2001). It is a complex process to interpret the correlations in the diagonals, triangles and blocks within a Multitrait-Multimethod Matrix

in order to assess the convergent and discriminant validities. The assessment of convergent and discriminant validity with this matrix could provide further evidence for construct validity of the instrument.

CME seeking requirements differs across countries. Nationality or resident country of the participating physicians for the future international CME activities needs to be recorded and considered to better interpret the study results.

Further research could be conducted to improve the measurement of some of the constructs. Two more items could be added to the subscales of positive beliefs, negative beliefs and behavioral intentions in order to increase the number of items per subscale. Cronbach's alpha coefficients for these subscales, therefore, could be more stable and accurate. The newly developed items according to other CME activities that focus on other clinical domains should be examined in future research.

The current study examined and established reliability and validity for the adapted CME evaluation instrument. These preliminary findings could serve as a research base for future TPB theory testing research on CME evaluation activities. A structural equation model could be built upon the validated inter-factor and factor-variable relationships to further confirm the underlying theoretical framework of this instrument as well as investigate the nuances among those relationships.

The current instrument development research could serve as an initial pilot for future CME evaluation studies. The value of this study was shown by the established validity and reliability of the adapted instrument, which provided credibility for the developed generic instrument template. However, the adapted instrument in current research focused on pre-operative breast cancer therapy only. In order to answer the second part

of question 5, future research needs to be conducted. This future research would adapt the developed template to other CME activities in order to examine whether the items and their corresponding constructs can be adapted to evaluate other CME conferences addressing different clinical domains. This future research could also help assure the consistency of the research results.

In order to help adapt the developed template to other CME evaluation activities in a rigorous way, future NIH workshops are encouraged to train CME meeting managers, organizers, and evaluators in developing the content specific CME evaluation instruments from the instrument template. Group process is suggested to develop a guidebook for CME evaluation instrument development. This guide book would include the developed instrument template and adaptation guidelines and be both available as hard copy and downloadable as electronic copy from NIH official website for CME evaluators, CME meeting managers and organizers to access. This guide book would provide a step-by-step guidance for other investigators who might not necessarily be social and behavioral scientists to help them adapt the template to their own medical domain.

In addition, evaluation design should include pretest, posttest, and follow-up data collection to further evaluate the ability of the instrument to measure changes in attitudes, beliefs, intentions, and, ultimately, the effectiveness of the CME intervention. Even better, behavioral intention in the instrument could be validated by assessing actual behavior—physicians' practice at the clinical setting. Such validations would bridge the Kirkpatrick's evaluation model and the Theory of Planned Behavior.

Conclusions

A valid, reliable, and adaptable instrument was developed for evaluating physicians' positive beliefs, negative beliefs, perceived behavioral control, subjective norms, attitudes, and behavioral intention. This was accomplished by providing evidence that supports major aspects of validity (e.g. content, substantive theories, structure, construct) and reliability (internal consistency, inter-item correlations, item-total correlations) for the instrument. Most of the findings and the information provided by experts suggested that the instrument measured characteristics it was intended to assess.

Summary

Findings from this study verified that a reliable instrument was adapted from the CME evaluation template to assess the physicians' positive beliefs, negative beliefs, attitudes, subjective norms and behavioral intentions regarding pre-operative breast cancer therapy. The researcher achieved the goals of this study, which were to: (1) conduct a thorough content validation process to satisfy the needs for instrument validity; (2) develop a CME evaluation template and a adapted instrument with subscales in theoretical domains based on TPB, i.e. positive belief, negative belief, attitude, subjective norm, perceived behavioral control and behavior intention; (3) examine the existence of unexpected underlying subscales of the draft instrument; (4) examine the psychometric properties of the draft instrument in order to investigate the acceptability of the levels of reliability for each of the necessary subscales; (5) investigate the appropriateness of the draft instrument through a thorough development process for CME evaluation. Future research is needed to examine its adaptability to other clinical domains in evaluating different CME activities.

APPENDIX A

Alternative Instrument Format A

1.	RECOMMENDING ***pre-operative systemic hormonal therapy*** to operable breast cancer patients is…	Harmful	Beneficial
2.	REFERRING appropriate operable breast cancer patients for ***pre-operative systemic hormonal therapy*** trials is…	Harmful	Beneficial
3.	RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients is…	Harmful	Beneficial
4.	REFERRING operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials is…	Harmful	Beneficial

	My colleagues think I should…		
5.	Refer operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials.	Unlikely	Likely
6.	Refer operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unlikely	Likely
7.	Recommend ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unlikely	Likely
8.	Recommend ***pre-operative hormonal therapy chemotherapy*** to operable breast cancer patients.	Unlikely	Likely

	Pre-operative (*as opposed to post-operative*) ***systemic chemotherapy*** will…		
9.	*improve the breast conservation rates* of operable breast cancer patients.	Unlikely	Likely
10.	*lead to a lower mortality rate* in the operable breast cancer patients.	Unlikely	Likely
11.	*reduce the overall medical costs* for operable breast cancer patients.	Unlikely	Likely
12.	have the *minimum side effects* for operable breast cancer patients.	Unlikely	Likely

	Rate your confidence level in…			
13.	*evaluating/assessing* operable breast cancer patients' suitability for receiving ***pre-operative systemic chemotherapy***.	Unconfident	☐☐☐☐☐☐☐	Confident
14.	*evaluating* operable breast cancer patients' suitability for ***pre-operative systemic hormonal therapy***.	Unconfident	☐☐☐☐☐☐☐	Confident
15.	*referring* operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unconfident	☐☐☐☐☐☐☐	Confident
16.	RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unconfident	☐☐☐☐☐☐☐	Confident
17.	RECOMMENDING ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.	Unconfident	☐☐☐☐☐☐☐	Confident

	I Intend to…			
18.	REFER operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials.	Unlikely	☐☐☐☐☐☐☐	Likely
19.	REFER operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unlikely	☐☐☐☐☐☐☐	Likely
20.	RECOMMEND ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unlikely	☐☐☐☐☐☐☐	Likely
21.	RECOMMEND ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.	Unlikely	☐☐☐☐☐☐☐	Likely

APPENDIX B

Alternative Instrument Format B

22	RECOMMENDING ***pre-operative systemic hormonal therapy*** to operable breast cancer patients is…	Harmful	1	2	3	4	5	6	7	Beneficial
23	REFERRING appropriate operable breast cancer patients for ***pre-operative systemic hormonal therapy*** trials is…	Harmful	1	2	3	4	5	6	7	Beneficial
24	RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients is…	Harmful	1	2	3	4	5	6	7	Beneficial
25	REFERRING operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials is…	Harmful	1	2	3	4	5	6	7	Beneficial

	My colleagues think I should…									
26	Refer operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials.	Unlikely	1	2	3	4	5	6	7	Likely
27	Refer operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unlikely	1	2	3	4	5	6	7	Likely
28	Recommend ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely
29	Recommend ***pre-operative hormonal therapy chemotherapy*** to operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely

	Pre-operative (*as opposed to post-operative*) ***systemic chemotherapy*** will…									
30.	*improve the breast conservation rates* of operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely
31.	*lead to a lower mortality rate* in the operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely
32.	*reduce the overall medical costs* for operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely
33.	have the *minimum side effects* for operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely

	Rate your confidence level in…									
34.	*evaluating/assessing* operable breast cancer patients' suitability for receiving ***pre-operative systemic chemotherapy***.	Unconfident	1	2	3	4	5	6	7	Confident
35.	*evaluating* operable breast cancer patients' suitability for ***pre-operative systemic hormonal therapy***.	Unconfident	1	2	3	4	5	6	7	Confident
36.	*referring* operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unconfident	1	2	3	4	5	6	7	Confident
37.	RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unconfident	1	2	3	4	5	6	7	Confident
38.	RECOMMENDING ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.	Unconfident	1	2	3	4	5	6	7	Confident

	I Intend to…									
39.	REFER operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials.	Unlikely	1	2	3	4	5	6	7	Likely
40.	REFER operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unlikely	1	2	3	4	5	6	7	Likely
41.	RECOMMEND ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely
42.	RECOMMEND ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7	Likely

APPENDIX C

Alternative Instrument Format C

TABLE 1.

(move the trial questions last)—this would be question 3 RECOMMENDDING ***pre-operative systemic hormonal therapy*** to operable breast cancer patients is

harmful	extremely	quite	slightly	neither	slightly	quite	extremely	**beneficial**

MY REFERRING appropriate operable breast cancer patients for ***pre-operative systemic hormonal therapy*** trials is this is question 4-hormonal therapy is 2 works

harmful	extremely	quite	slightly	neither	slightly	quite	extremely	**beneficial**

MY RECOMMENDDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients is

harmful	extremely	quite	slightly	neither	slightly	quite	extremely	**beneficial**

MY REFERRING operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials is

harmful	extremely	quite	slightly	neither	slightly	quite	extremely	**beneficial**

TABLE 2

MY COLLEAGUES THINK I should refer operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

MY COLLEAGUES THINK I should recommend ***pre-operative systemic chemotherapy*** to operable breast cancer patients.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

MY COLLEAGUES THINK I should recommend ***pre-operative hormonal therapy chemotherapy*** to operable breast cancer patients.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

Please indicate who influence your clinical decisions most_____

Pre-operative *(as opposed to post-operative)* ***systemic chemotherapy*** will have the *minimum side effects* for operable breast cancer patients.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

Pre-operative *(as opposed to post-operative)* ***systemic chemotherapy*** will *improve the breast conservation rates* of operable breast cancer patients.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

Pre-operative *(as opposed to post-operative)* ***systemic chemotherapy*** will *lead to a lower mortality rate* in the operable breast cancer patients.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

Pre-operative *(as opposed to post-operative)* ***systemic chemotherapy*** will *reduce the overall medical costs* for operable breast cancer patients.

unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**

TABLE 4. This section asks for your evaluation of beliefs that the ***Pre-operative*** (*as opposed to post-operative*) ***systemic chemotherapy*** will produce a given outcome.								
Pre-operative (*as opposed to post-operative*) ***systemic chemotherapy*** *improving the breast conservation rates* of operable breast cancer patients is								
bad	extremely	quite	slightly	neither	slightly	quite	extremely	**good**
omit this question. Sounds "silly" as why would lower mortality rates be bad								
bad	extremely	quite	slightly	neither	slightly	quite	extremely	**good**
bad	extremely	quite	slightly	neither	slightly	quite	extremely	**good**
bad	extremely	quite	slightly	neither	slightly	quite	extremely	**good**

TABLE 6. This section asks for your readiness to engage the ***Pre-operative*** (*as opposed to post-operative*) ***systemic chemotherapy.***								
I INTEND TO REFER operable breast cancer patients to appropriate ***pre-operative systemic chemotherapy*** trials.								
TABLE 5.								
unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**
Rate your confidence level in *evaluating/assessing* operable breast cancer patients' suitability for								
I INTEND TO REFER operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.								
unconfident	extremely	quite	slightly	neither	slightly	quite	extremely	**confident**
unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**
I INTEND TO RECOMMEND ***pre-operative systemic chemotherapy*** to operable breast cancer patients.								
unconfident	extremely	quite	slightly	neither	slightly	quite	extremely	**confident**
unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**
I INTEND TO RECOMMEND ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.								
unlikely	extremely	quite	slightly	neither	slightly	quite	extremely	**likely**
unconfident	extremely	quite	slightly	neither	slightly	quite	extremely	**confident**
Rate your confidence level of *referring* operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.								
unconfident	extremely	quite	slightly	neither	slightly	quite	extremely	**confident**
Rate your confidence level in RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients.								
unconfident	extremely	quite	slightly	neither	slightly	quite	extremely	**confident**
Rate your confidence level in RECOMMENDING ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.								
unconfident	extremely	quite	slightly	neither	slightly	quite	extremely	**confident**

APPENDIX D

Continuing Medical Education Evaluation Survey (draft)

This questionnaire is designed to help the NIH CME office evaluate the effectiveness of the conference entitled: *Preoperative Therapy in Invasive Breast Cancer*. There is evidence to support the use of post-operative systemic chemotherapy in a subset of women with operative breast cancer. This survey focuses on the use of ***pre-operative*** **systemic chemotherapy** in this population. Thank you for your assistance.

Please circle your specialty	Medical Oncologist	Surgeon	Pathologist	Radiation Oncologist	
	Radiologist	Physician	Registered Nurse	Other(Please specify)	____
Please circle your affiliation	Academia		Government	Industry	
	Community Practice		Other(Please specify)	_____	
Please indicate number of years in practice:					
Your date of birth (mm/dd/yy):					
Your initials:					
Are you seeking CME credits?			YES	NO	
Your gender:			FEMALE	MALE	

	Pre-operative (as opposed to post-operative) systemic chemotherapy will…								
43.	improve breast conservation rates of operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
44.	increase local recurrence rates of operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
45.	increase disease-free survival rates for operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
46.	increase the risk of inadequate surgery for operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
47.	lead to a lower mortality rate in operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
48.	reduce the overall medical costs for operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
49.	have fewer side effects for operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
	Please rate your confidence level in…								
50.	SHARING information about ***pre-operative systemic chemotherapy*** with operable breast cancer **patients**.	Unconfident	1	2	3	4	5	6	7 Confident
51.	SHARING knowledge of ***pre-operative systemic chemotherapy*** with **physicians** who do not attend the conference.	Unconfident	1	2	3	4	5	6	7 Confident
52.	EVALUATING/ASSESSING operable breast cancer patients' suitability for receiving ***pre-operative systemic chemotherapy***.	Unconfident	1	2	3	4	5	6	7 Confident
53.	RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unconfident	1	2	3	4	5	6	7 Confident
54.	REFERRING operable breast cancer patients to ***pre-operative systemic chemotherapy*** trials.	Unconfident	1	2	3	4	5	6	7 Confident
55.	APPLYING knowledge of ***pre-operative systemic chemotherapy*** in developing research studies.	Unconfident	1	2	3	4	5	6	7 Confident
56.	EVALUATING ***pre-operative systemic chemotherapy papers*** critically when they appear in the literature.	Unconfident	1	2	3	4	5	6	7 Confident

SURVEY CONTINUES IN THE BACK

The practice of pre-operative systemic chemotherapy is…									
57.	Not credible	1	2	3	4	5	6	7	Credible
58.	Unsafe	1	2	3	4	5	6	7	Safe
59.	Harmful	1	2	3	4	5	6	7	Beneficial
60.	Ineffective	1	2	3	4	5	6	7	Effective
61.	Frustrating	1	2	3	4	5	6	7	Satisfying
62.	Impractical	1	2	3	4	5	6	7	Useful
63.	Hard	1	2	3	4	5	6	7	Easy

My colleagues think I should…									
64.	SHARE information about ***pre-operative systemic chemotherapy*** with operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
65.	SHARE knowledge of ***pre-operative systemic chemotherapy*** with physicians who do not attend the conference.	Unlikely	1	2	3	4	5	6	7 Likely
66.	RECOMMEND ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
67.	REFER operable breast cancer patients to ***pre-operative systemic chemotherapy*** trials.	Unlikely	1	2	3	4	5	6	7 Likely

Please indicate who influence your clinical decision most: colleagues supervisor patients other (please specify) _____

	I Intend to…								
68.	SHARE information about ***pre-operative systemic chemotherapy*** with operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely
69.	SHARE knowledge of ***pre-operative systemic chemotherapy*** with physicians who do not attend the conference.	Unlikely	1	2	3	4	5	6	7 Likely
70.	REVIEW THE LITERATURE about ***pre-operative systemic chemotherapy***.	Unlikely	1	2	3	4	5	6	7 Likely
71.	APPLY knowledge of ***pre-operative systemic chemotherapy*** in developing research studies.	Unlikely	1	2	3	4	5	6	7 Likely
72.	EVALUATE ***pre-operative systemic chemotherapy*** critically when they appear in the literature.	Unlikely	1	2	3	4	5	6	7 Likely

*The last two questions examine the use of pre-operative **hormonal** therapy in breast cancer patients eligible for post-operative systemic chemotherapy.*

73.	REFER operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	Unlikely	1	2	3	4	5	6	7 Likely
74.	RECOMMEND ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.	Unlikely	1	2	3	4	5	6	7 Likely

THANK YOU

APPENDIX E

Cognitive Testing Instruction

1. Introduction

We are testing a questionnaire today. The questionnaire is in a draft format now. After it is finalized, it will be given to the participating physicians in the conference entitled: P*reoperative Therapy in Invasive Breast Cancer* sponsored by NCI on March 26-27.

Before we finalize it, we want to know if any of the questions are difficult to understand, hard to answer, or do not make sense. That's why we asking you to try it out for us and tell us what you think as you go along.

Don't worry about making any criticisms about the questionnaire. You won't hurt anyone's feelings. We just need your honest comments.

2. Practicing Think-Aloud

While you are going through the questionnaire I am going to ask you to think aloud so that I can understand if there are any problems with the questionnaire. By "think aloud", I mean reading all the questions aloud and telling me what you are thinking as you read the questions and as you pick your answers. The first thing we will do is practice thinking aloud.

Try to read the questions aloud to yourself. Answer them by circling the number on the questionnaire while tell me what was the number selected and why it was selected.

3. Questionnaire

APPENDIX F

IRB Application

Instrument Development for Continuing Medical Education (CME) Evaluation

1. Abstract

The purpose of this project is to develop a standardized, theory-based, valid and reliable pre/post activity evaluation instrument for clinicians with core items assessing the constructs of the Theory of Planned Behavior. The draft questionnaire was created by developing core questions to address the constructs then adapting them according to the learning objectives of one NIH CME conference. There are 750 physicians anticipated to attend the conference, and both pre-test and post-test data will be collected. Exploratory factor analysis will be conducted with the pretest data to examine the structure of the instrument and potential subscales. Item analysis will be conducted to examine the internal consistency reliability of any subscales that emerge. Convergent and discriminant validity will also be examined for the subscales. Post-test data will be analyzed to assess attitude and behavioral intention change related to preoperative therapies for breast cancer.

Informed consent will be obtained from all participants and all responses will be confidential. This IRB submission concerns the data collection, analysis and CME evaluation survey. Use of a standardized instrument will enable comparison of effectiveness across different CME interventions, helping researchers understand factors influencing the effectiveness of different CME programs and guiding future CME intervention and evaluation design.

2. Subject Selection

a. Who will be the subjects? How will you enlist their participation? If you plan to advertise for subjects, please include a copy of the advertisement.

The instrument is going to be administered to the participants of the meeting of Preoperative Therapy in Invasive Breast Cancer: Reviewing the State of the Science and Exploring New Research Directions to be held at March 26 and 27, 2007, in the Natcher Conference Center, National Institutes of Health, Bethesda, Maryland. The clinical domain of items will be the one addressed in this conference, e.g. preoperative therapy in invasive breast cancer. The meeting purpose and the conference objectives were used to operationalize adaptable measures for CME activities designed to address physician practices.

There are 750 physicians anticipated to attend the conference, and the questionnaires (Appendix A) will be included in the advance registration package sent to the 750 physician registrants, who will be asked to return it at the beginning of the two-day conference. Post-test questionnaires will be administered at lunch time on the second day of the conference. The meeting organizer will mention the questionnaire briefly at the beginning of the two day conference to facilitate the return of the questionnaire. All the participating physicians attending the conference and who agree to participate will be included in the sample.

b. Will the subjects be selected for any specific characteristics (e.g., age, sex, race, ethnic origin, religion, or any social or economic qualifications)?

There is no selection criteria for the participants based on any specific characteristics (e.g., age, sex, race, ethnic origin, religion, or any social or economic qualifications). Attendees who are not physicians will be excluded from the analysis.

c. State why the selection will be made on the basis or bases given in 2(b).

The focus of the study is physician attitudes and behavior.

3. Procedures

The proposed paper-pencil instruments (Appendix A) will be attached to the required NIH CME evaluation form and administered in the conference package upon registration. Participating physicians will be asked to return the evaluation instruments at the beginning of the two-day conference. The pre-test data collected will be used to develop the CME instrument. The consent forms (Appendix B) will be attached to and administered along with the pre-test instruments. It will describe the nature of the evaluation survey, cooperation requested from the participants and assured privacy and confidentiality for the participants. The consent forms will be collected along with the pre-test instruments. Post-test questionnaires will be administered at lunch time of the second day of the conference.

Several strategies will be used to protect human subjects. Informed consent will be obtained from all participating physicians and all responses will be confidential. Pretest data will be collected before the conference. Information will be collected under confidentiality, e.g. conference registration numbers will be used as the identifier so that participants' information could not be identified by the principle investigator (PI) either directly or through identifiers. Data will be reported in aggregate form thus individual identification will not be tied to data analysis and reporting. Conference registration numbers will be used to link the pre-test and post-test data. Pre-test and post-test data will be compared to examine the effectiveness of this CME conference. All project staffs have been trained in confidentiality procedures.

4. Risks and Benefits

There will be no physical, social, or legal risks of any kind to the participants. Risks to study participants are minimal. Response to the questions in the scale is not expected to cause discomfort or anxiety among participants. The project is designed to help develop the evaluation instrument as well as assess the effectiveness of the conference intervention. Individual participants' attitudes and behavioral intentions will not be the focus of this study but only used to evaluate the intervention.

The participants might not benefit directly from participating in the project and filling out the survey. However, the information collected from this project will help the development of a valid, reliable, and adaptable evaluation instrument for NIH to use in its future

conferences. Hopefully, future NIH CME conference instructions and evaluations will benefit from this instrument. As the scale questions will be integrated with the official NIH CME evaluation form, participants will be encouraged to fill out and return the survey to the most extent.

5. Confidentiality

Data will be collected before the conference. Information will be collected under confidentiality, e.g. conference registration numbers will be used as the identifier that participants' information could not be identified by the principle investigator (PI) either directly or through identifiers. Data will be reported in aggregate form thus individual identification will not be tied to data analysis and reporting.

Completed surveys will be collected by conference instructors and placed in a sealed envelope. Surveys will be removed only by the NIH CME committee members or UMD researchers. Data from the survey will be coded for easy analyzing, interpreting, and reporting. Surveys will be kept at the Public Health Informatics Research Laboratory at the University of Maryland in a locked file cabinet. Only CME committee members and project researchers at the University of Maryland will have the access to them. Surveys will be returned to NIH CME office to be shredded upon completion of the research.

6. Information and Consent Forms

The consent forms (Appendix B) will be administered along with the pre-test instruments. It will describe the nature of the evaluation survey, cooperation requested from the participants and assured privacy and confidentiality for the participants.

7. Conflict of Interest

There is no conflict of interest.

8. HIPAA Compliance

This project will not use protected health information.

9. Research Outside of the United States:

Not applicable

10. Research Involving Prisoners:

Not applicable

11. Appendices

A. CME evaluation questionnaire

B. The consent forms

Initials _____ Date _____

CONSENT FORM

Project Title	Instrument Development/Program Evaluation for Continuing Medical Education (CME) Evaluation
Why is this research being done?	This is a research project being conducted by Jing Tian at the University of Maryland, College Park. We are inviting you to participate in this research project because you are a participating physician of CME conference The purpose of this research project is to collect pretest and posttest questionnaire data in order to develop the CME evaluating instrument as well as evaluate the effectiveness of this CME conference.
What will I be asked to do?	A paper-pencil instrument will be attached to the required NIH CME evaluation form and administered in the conference package upon registration. You will be asked to return the evaluation instruments at the beginning of the two-day conference. Post-test questionnaires will be administered at the lunch time of the second day of the conference.
What about confidentiality?	We will do our best to keep your personal information confidential Data will be collected before the conference. Information will be collected under confidentiality, e.g. conference registration numbers will be used as the identifier that participants' information could not be identified by the principle investigator (PI) either directly or through identifiers. Data will be reported in aggregate form thus individual identification will not be tied to data analysis and reporting. Completed surveys will be collected by conference instructors and placed in a sealed envelope. Surveys will be removed only by the NIH CME committee members or UMD researchers. Data from the survey will be coded for easy analyzing, interpreting, and reporting. Surveys will be kept at the Public Health Informatics Research Laboratory at the University of Maryland in a locked file cabinet. Only CME committee members and project researchers at the University of Maryland will have the access to them. Surveys will be returned to NIH CME office to be shredded upon completion of the research.
What are the risks of this research?	There will be no physical, social, or legal risks of any kind to the participants. Risks to study participants are minimal. Response to the questions in the scale would not cause discomfort or anxiety among participants. The project is designed to help develop the evaluation instrument as well as assess the effectiveness of the conference intervention. Individual participants' attitudes, behavioral beliefs, evaluation of behavioral beliefs, subjective norms, perceived behavior control and behavioral intentions will not be the focus of this study but only used to evaluate the intervention.

Initials _____ Date _____

Project Title	Instrument Development/Program Evaluation for Continuing Medical Education (CME) Evaluation
What are the benefits of this research?	The participants might not benefit directly from participating in the project and filling out the survey. However, the information collected from this project will help the development of a valid, reliable, and adaptable evaluation instrument for NIH to use in its future conferences. Hopefully, future NIH CME conference instructions and evaluations will benefit from this instrument. As the scale questions will be integrated with the official NIH CME evaluation form, participants will be encouraged to fill out and return the survey to the most extent.
Do I have to be in this research? **Can I stop participating at any time?**	Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.
Is any medical treatment available if I am injured?	The University of Maryland does not provide any medical, hospitalization or other insurance for participants in this research study, nor will the University of Maryland provide any medical treatment or compensation for any injury sustained as a result of participation in this research study, except as required by law.
What if I have questions?	This research is being conducted by Jing Tian at the University of Maryland, College Park. If you have any questions about the research study itself, please contact Jing Tian at: The University of Maryland, Suite 2387 Valley Drive, HHP, 301-405-9626 or tianjing@umd.edu If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, Maryland, 20742; (e-mail) irb@deans.umd.edu; (telephone) 301-405-0678 This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.
Statement of Age of Subject and Consent	Your signature indicates that: you are at least 18 years of age;, the research has been explained to you; your questions have been answered; and you freely and voluntarily choose to participate in this research project.

Project Title	Instrument Development/Program Evaluation for Continuing Medical Education (CME) Evaluation	
Signature and Date	**NAME OF SUBJECT**	
	SIGNATURE OF SUBJECT	
	DATE	

******Please note: When consent form requires more than one page, please include a space for the subject to initial and date at the top right-hand corner of each page. The corner should appear as: Initials_____ Date_____**
Also, each page must display a page range such as: Page 1 of 2, then Page 2 of 2. This step would confirm that the subject agreed to the entire contents of the consent form. ****

APPENDIX G

Continuing Medical Education Evaluation Survey (Pre-test)

This questionnaire is designed to help the NIH CME Office evaluate the effectiveness of the conference entitled: *Preoperative Therapy in Invasive Breast Cancer*. There is evidence to support the use of post-operative systemic chemotherapy in a subset of women with operative breast cancer. This survey focuses on the use of ***pre-operative* systemic chemotherapy** in this population. Thank you for your collaboration.

Please circle your specialty	Medical Oncologist	Surgeon	Pathologist	Radiation Oncologist
	Radiologist	Registered Nurse	Retired	Other (Please specify) _____
Please circle your dominant affiliation	Academia	Government	Industry	
	Community Practice	Other (Please specify) _____		
Please indicate number of years in	Patient care ONLY: _____	Research (non-patient care) ONLY: _____		
	Patient care AND Research at the same time: _____			
Your age:				
Do you seek CME credits?	YES	NO		
Your gender:	FEMALE	MALE		

Pre-operative (as opposed to post-operative) systemic chemotherapy will:	**Unlikely**						**Likely**
75. improve breast conservation rates of operable breast cancer patients.	1	2	3	4	5	6	7
76. increase local recurrence rates of operable breast cancer patients.	1	2	3	4	5	6	7
77. increase disease-free survival rates for operable breast cancer patients.	1	2	3	4	5	6	7
78. increase the risk of inadequate surgery for operable breast cancer patients.	1	2	3	4	5	6	7
79. lead to a lower mortality rate in operable breast cancer patients.	1	2	3	4	5	6	7
80. reduce the overall medical costs for operable breast cancer patients.	1	2	3	4	5	6	7
81. have fewer side effects for operable breast cancer patients.	1	2	3	4	5	6	7

Please rate your confidence level in…	**Not confident**						**Confident**
82. SHARING information about ***pre-operative systemic chemotherapy*** with operable breast cancer **patients**.	1	2	3	4	5	6	7
83. SHARING knowledge of ***pre-operative systemic chemotherapy*** with **physicians** who do not attend the conference.	1	2	3	4	5	6	7
84. EVALUATING/ASSESSING operable breast cancer patients' suitability for receiving ***pre-operative systemic chemotherapy***.	1	2	3	4	5	6	7
85. RECOMMENDING ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	1	2	3	4	5	6	7
86. REFERRING operable breast cancer patients to ***pre-operative systemic chemotherapy*** trials.	1	2	3	4	5	6	7
87. APPLYING knowledge of ***pre-operative systemic chemotherapy*** in developing or deciding to participate in research studies as a researcher.	1	2	3	4	5	6	7
88. EVALUATING ***pre-operative systemic chemotherapy papers*** critically when they appear in the literature.	1	2	3	4	5	6	7

SURVEY CONTINUES ON THE BACK

The practice of pre-operative (as opposed to post-operative) systemic chemotherapy is… (Please circle the number)									
89.	Not credible	1	2	3	4	5	6	7	Credible
90.	Unsafe	1	2	3	4	5	6	7	Safe
91.	Harmful	1	2	3	4	5	6	7	Beneficial
92.	Ineffective	1	2	3	4	5	6	7	Effective
93.	Frustrating	1	2	3	4	5	6	7	Satisfying
94.	Impractical	1	2	3	4	5	6	7	Useful
95.	Complex	1	2	3	4	5	6	7	Simple

Please indicate who (select one) influences your decision-making most: Experts Peer Colleagues Senior Colleagues Patients Other(Please specify) _____

Most clinicians whose opinion I value think I should…	**Unlikely**						**Likely**
96. SHARE information about ***pre-operative systemic chemotherapy*** with operable breast cancer patients.	1	2	3	4	5	6	7
97. SHARE knowledge of ***pre-operative systemic chemotherapy*** with physicians who do not attend the conference.	1	2	3	4	5	6	7
98. RECOMMEND ***pre-operative systemic chemotherapy*** to operable breast cancer patients.	1	2	3	4	5	6	7
99. REFER operable breast cancer patients to ***pre-operative systemic chemotherapy*** trials.	1	2	3	4	5	6	7

I Intend to…	**Unlikely**						**Likely**
100. SHARE information about ***pre-operative systemic chemotherapy*** with operable breast cancer patients.	1	2	3	4	5	6	7
101. SHARE knowledge of ***pre-operative systemic chemotherapy*** with physicians who do not attend the conference.	1	2	3	4	5	6	7
102. REVIEW LITERATURE about ***pre-operative systemic chemotherapy*** in the next month.	1	2	3	4	5	6	7
103. APPLY knowledge of ***pre-operative systemic chemotherapy*** in developing or deciding to participate in research studies as a researcher.	1	2	3	4	5	6	7
104. EVALUATE ***pre-operative systemic chemotherapy papers*** critically when they appear in the literature.	1	2	3	4	5	6	7

*The last two items examine the use of pre-operative **hormonal** therapy in breast cancer patients eligible for post-operative systemic chemotherapy.*

I Intend to…	**Unlikely**						**Likely**
105. REFER operable breast cancer patients to appropriate ***pre-operative systemic hormonal therapy*** trials.	1	2	3	4	5	6	7
106. RECOMMEND appropriate ***pre-operative systemic hormonal therapy*** to operable breast cancer patients.	1	2	3	4	5	6	7

THANK YOU

REFERENCES

Agrawal S, Saluja I, Kaczorowski J. (2004). A prospective before-and-after trial of an educational intervention about pharmaceutical marketing. *Acad Med*, 79 (11): 1046-1050.

Aguilar R, Garay J, Villatoro M, Billo NE, Caminero JA. (2004). Impact of a model training course for private and public specialist physicians in El Salvador. *Int J Tuberc Lung Dis*, 8(4):473-479.

Alexander JH, Peterson ED, Chen AY, Harding TM, Adams DB, Kisslo JA Jr. (2004). Feasibility of point-of-care echocardiography by internal medicine house staff. *Am Heart J*; 147(3):476-481.

Allard J, Hebert R, Rioux M, Asselin J & Voyer L. (2001). Efficacy of a clinical medication review on the number of potentially inappropriate prescriptions prescribed for community-dwelling elderly people. *CMAJ*, 164(9): 1291-1296.

Allen M, Sargeant J, Mann K, Fleming M, Premi J. (2003). Videoconferencing for practice-based small-group continuing medical education: feasibility, acceptability, effectiveness, and cost. *J Contin Educ Health Prof*, 23(1):38-47.

American Medical Association. (2006). State Medical Licensure Requirements and Statistics. Maryland: Data Trace.

Backhaus J, Junghanns K, Mueller-Popkes K, Broocks A, Riemann D, Hajak G, Hohagen F. (2002). Short-term training increases diagnostic and treatment rate for insomnia in general practice. *Eur Arch Psychiatry Clin Neurosci*, 252(3):99-104.

Bellamy N, Goldstein LD, Tekanoff RA. (2000). Continuing medical education-driven skills acquisition and impact on improved patient outcomes in family practice setting. *J Contin Educ Health Prof,* 20(1):52-61.

Ben-Abraham R, Weinbroum AA, Kluger Y, Stein M, Barzilay Z, Paret G. (2000). Pediatricians and the Advanced Trauma Life Support (ATLS): time for reconsideration. *Isr Med Assoc J*, 2(7):513-516.

Ben-Arye E, Frenkel M. (2004). An approach to teaching physicians about complementary medicine in the treatment of cancer. *Integr Cancer Ther*, 3(3):208-213.

Bloom BS. (2005). International Journal of Technology Assessment in Health Care Cambridge University Press, 21: 380-385.

Bland DR, Dugan E, Cohen SJ, Preisser J, Davis CC, McGann PE, Suggs PK, Pearce KF. (2003). The effects of implementation of the Agency for Health Care Policy and Research urinary incontinence guidelines in primary care practices. *J Am Geriatr Soc*, 51(7):979-984.

Boakes J, Gardner D, Yuen K, Doyle S. (2000). General practitioner training in palliative care: an experiential approach. *J Palliat Care*, 16(2):11-19.

Botash AS, Galloway AE, Booth T, Ploutz-Snyder R, Hoffman-Rosenfeld J, Cahill L. (2005) Continuing medical education in child sexual abuse: cognitive gains but not expertise. *Arch Pediatr Adolesc Med*, 159(6):561-566.

Bozeman WP, Eastman ER. (2002). Tactical EMS: an emerging opportunity in graduate medical education. *Prehosp Emerg Care*, 6(3):322-324.

Bracchi RC, Houghton J, Woods FJ, Thomas S, Smail SA, Routledge PA. (2005). A distance-learning programme in pharmacovigilance linked to educational credits is associated with improved reporting of suspected adverse drug reactions via the UK yellow card scheme. *Br J Clin Pharmacol*, 60(2):221-223.

Broers S, Smets E, Bindels P, Evertsz' FB, Calff M, de Haes H. (2005). Training general practitioners in behavior change counseling to improve asthma medication adherence *Patient Educ Couns*, 58(3):279-287.

Buckley LL, Goering P, Parikh SV, Butterill D, Foo EK. (2003). Applying a 'stages of change' model to enhance a traditional evaluation of a research transfer course. *J Eval Clin Pract*, 9(4):385-90.

Byszewski AM, Graham ID, Amos S, Man-Son-Hing M, Dalziel WB, Marshall S, Hunt L, Bush C. (2003). A continuing medical education initiative for canadian primary care physicians: the driving and dementia toolkit: a pre- and postevaluation of knowledge, confidence gained, and satisfaction. *J Am Geriatr Soc*, 51(10):1484-1489.

Carli P, De Giorgi V, Crocetti E, Caldini L, Ressel C, Giannotti B. (2005). Diagnostic and referral accuracy of family doctors in melanoma screening: effect of a short formal training. *Eur J Cancer Prev*, 14(1):51-55.

Casebeer L, Kristofco RE, Strasser S, Reilly M, Krishnamoorthy P, Rabin A, Zheng S, Karp S. (2004). Standardizing evaluation of on-line continuing medical education: physician knowledge, attitudes, and reflection on practice. *J Contin Educ Health Prof,* 24(2):68-75.

Casebeer LL, Strasser SM, Spettell CM, Wall TC, Weissman N, Ray MN. Allison JJ. (2003). Designing tailored Web-based instruction to improve practicing physicians' preventive practices. *J Med Internet Res,* 5(3):e20.

Chan CS, Wun YT, Cheung A, Dickinson JA, Chan KW, Lee HC, Yung YM. (2003). Communication skill of general practitioners: any room for improvement? How much can it be improved? *Med Educ,* 37(6):514-526.

Chandalia HB, Purohit S, Lamba PS. (2000). Evaluation of a short course on diabetology for practising doctors. *J Assoc Physicians India*, 48(11):1091-1095.

Cheng GY. (2004). A study of clinical questions posed by hospital clinicians. *J Med Libr Assoc*, 92(4): 445-458.

Clever SL, Novack DH, Cohen DG, Levinson W. (2003). Evaluating surgeons' informed decision making skills: pilot test using a videoconferenced standardized patient. *Med Educ*, 37(12):1094-1099.

Cole TB, Glass RM. (2004). Learning associated with participation in journal-based continuing medical education. *J Contin Educ Health Prof*, 24(4):205-212.

Confos N, Frith J, Mitchell P. (2003). Training GPs to screen for diabetic retinopathy. The impact of short term intensive education. *Aust Fam Physician*, 32(5):381-382, 384.

Curran VR, Fleet L. (2005). A review of evaluation outcomes of web-based continuing medical education. *Medical Education,* 39: 561-567.

Curran VR, Hoekman T, Gulliver W, Landells I, Hatcher L. (2000). Web-based continuing medical education. (II): Evaluation study of computer-mediated continuing medical education *J Contin Educ Health Prof*, 20(2):106-119.

Cutts C. (2001). Use of a clinical audit to promote the quality use of medicines to rural and remote general practitioners. *Aust J Rural Health*, 9 Suppl 1:S49-53.

Curtis P, Carey TS, Evans P, Rowane MP, Mills GJ, Jackman A. (2000). Training primary care physicians to give limited manual therapy for low back pain: patient outcomes. *Spine*, 25:2954-2960; discussion 2960-2961.

David AD, Mazmanian PE, Fordis M, Harrison RV, Thorpe KE, Perrier L. (2006). Accuracy of Physician Self-assessment Compared With Observed Measures of Competence A Systematic Review *JAMA*, 296:1094-1102.

Davis AD, Thomson O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. (1999). Impact of formal continuing medical education. *JAMA*, 282 (9): 867-874.

Davis P, Kvern B, Donen N, Andrews E, Nixon O. (2000). Evaluation of a problem-based learning workshop using pre- and post-test objective structured clinical examinations and standardized patients. *J Contin Educ Health Prof*, 20(3):164-170.

Davis RS, Bukstein DA, Luskin AT, Kailin JA, Goodenow G. (2004). Changing physician prescribing patterns through problem-based learning: an interactive, teleconference case-based education program and review of problem-based learning. *Ann Allergy Asthma Immunol*; 93(3):237-242.

Delvaux N, Merckaert I, Marchal S, Libert Y, Conradt S, Boniver J, Etienne AM, Fontaine O. (2005). Physicians' communication with a cancer patient and a relative: a randomized study assessing the efficacy of consolidation workshops. *Cancer*, 103, 2397-2411.

Derebery VJ, Giang GM, Saracino G, Fogarty WT. (2002). Evaluation of the impact of a low back pain educational intervention on physicians' practice patterns and patients' outcomes. *J Occup Environ Med*, 44(10):977-984.

de Villiers M, Bresick G, Mash B. (2003). The value of small group learning: an evaluation of an innovative CPD program for primary care medical practitioners. *Med Educ*, 37(9):815-821.

Dickinson JA, Chan CS, Wun YT, Tsang KW. (2002). Graduates' evaluation of a postgraduate diploma course in family medicine. *Fam Pract*, 19(4):416-421.

Dowell J, Pagliari C, McAleer S. (2004). Development and evaluation of a concordance training course for medical practitioners. *Med Teach*, 26(4): 384-386.

Evans A, Ali S, Singleton C, Nolan P, Bahrami J. (2002). The effectiveness of personal education plans in continuing professional development: an evaluation. *Med Teach*, 24(1):79-84.

Faber E, Bierma-Zeinstra SM, Burdorf A, Nauta AP, Hulshof CT, Overzier PM, Miedema HS, Koes BW. (2005). In a controlled trial training general practitioners and occupational physicians to collaborate did not influence sickleave of patients with low back pain. *J Clin Epidemiol*, 58(1):75-82.

Fabrizio MD, Tuerk I, Schellhammer PF. (2003). Laparoscopic radical prostatectomy: decreasing the learning curve using a mentor initiated approach. *J Urol*, 169(6): 2063-2065.

Filippucci E, Unlu Z, Farina A, Grassi W. (2003). Sonographic training in rheumatology: a self teaching approach. *Ann Rheum Dis*, 62(6):565-567.

Flores R, Robles J, Burkhalter BR. (2002). Distance education with tutoring improves diarrhea case management in Guatemala. *Int J Qual Health Care*, 14 Suppl 1:47-56.

Frey K, Edwards F, Altman K, Spahr N, Gorman RS. (2003). The 'Collaborative Care' curriculum: an educational model addressing key ACGME core competencies in primary care residency training. *Med Educ*, 37(9):786-789.

Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, Andrew CG. (2004). Proving the value of simulation in laparoscopic surgery. *Ann Surg*, 240(3):518-525; discussion 525-528.

Gask L, Dowrick C, Dixon C, Sutton C, Perry R, Torgerson D, Usherwood T. (2004). A pragmatic cluster randomized controlled trial of an educational intervention for GPs in the assessment and management of depression. *Psychol Med*, 34(1):63-72.

Gerbert B, Bronstone A, Maurer T, Berger T, McPhee SJ, Caspers N. (2002). The effectiveness of an Internet-based tutorial in improving primary care physicians' skin cancer triage skills. *J Cancer Educ*, 17(1):7-11.

Gielen AC, Wilson ME, McDonald EM, Serwint JR, Andrews JS, Hwang WT, Wang MC. (2001). Randomized trial of enhanced anticipatory guidance for injury prevention. *Arch Pediatr Adolesc Med,* 155(1):42-9.

Glazebrook R, Manahan D, Chater AB. (2005). Evaluation of an ultrasound program (intermediate obstetric and emergency medicine) for Australian rural and remote doctors. *Aust J Rural Health*, 13(5):295-299.

Goldberg HI, Deyo RA, Taylor VM, Cheadle AD, Conrad DA, Loeser JD, Heagerty PJ, Diehr P. (2001). Can evidence change the rate of back surgery? A randomized trial of community-based education. *Eff Clin Pract*, 4(3):95-104.

Gonsalves WC, Skelton J, Smith T, Hardison D, Ferretti G. (2004). Physicians' oral health education in Kentucky. *Fam Med*, 36(8): 544-546.

Gormley GJ, Steele WK, Stevenson M, McKane R, Ryans I, Cairns AP, Pendleton A, Wright GD. (2003). A randomized study of two training programs for general practitioners in the techniques of shoulder injection. *Ann Rheum Dis*, 62(10):1006-1009.

Guiscafre H, Gutierrez G, Verver H, Palafox M, Lopez A, Martinez H. (2003). Quality improvement of integrated child health care management after in-service training for physicians. *Fam Pract*; 20(1):74-76.

Haas LJ, Houchins J, Leiser JP. (2003). Changing family physicians' visit structuring behavior: a pilot study. *Fam Med*; 35(10):726-729.

Habermann R, Butts N, Powers J, Pichert JW. (2002). Physician behavior changes following CME on the prospective payment system in long-term care: a pilot study. *J Am Med Dir Assoc*; 3(1):12-15.

Hall L, Eccles M, Barton R, Steen N, Campbell M. (2001). Is untargeted outreach visiting in primary care effective? A pragmatic randomized controlled trial. *J Public Health Med*, 23(2):109-113.

Harris JM, Salasche SJ, Harris RB. (2001). Can Internet-based continuing medical education improve physicians' skin cancer knowledge and skills? *J Gen Intern Med*; 16(1):50-56.

Haug, T.T., Hellstrom, K., Blomhoff, S., Humble, M., Madsbu, H.P., Wold, J.E. et al. (2000). The treatment of social phobia in general practice. Is exposure therapy feasible? *Fam Pract*, 17(2): 114-118.

Heinemann AW, Roth EJ, Rychlik K, Pe K, King C, Clumpner J. (2003), The impact of stroke practice guidelines on knowledge and practice patterns of acute care health professionals. *J Eval Clin Pract*, 9(2):203-212.

Henderson DC, Mollica RF, Tor S, Lavelle J, Culhane MA, Hayden D. (2005). Building primary care practitioners' attitudes and confidence in mental health skills in a post-conflict society: a Cambodian example. *J Nerv Ment Dis*, 193(8):551-559.

Holroyd BR, Wilson D, Rowe BH, Mayes DC, Noseworthy T. (2004). Uptake of validated clinical practice guidelines: experience with implementing the Ottawa Ankle Rules. *Am J Emerg Med*, 22(3):149-155.

Jackson C, De Jong I, Schuluter PJ. Changing clinician practice. (2002). The RACGP/RACO National GP Eye Skills Workshop. *Aust Fam Physician*; 31(3):285-290.

Jacobs LM, Burns KJ, Luk SS, Marshall WT. (2005). Follow-up survey of participants attending the Advanced Trauma Operative Management (ATOM) Course. *J Trauma*, 58(6):1140-1143.

Jansen-van der Weide MC, Onwuteaka-Philipsen BD, van der Wal G. (2004). Implementation of the project 'Support and Consultation on Euthanasia in The Netherlands' (SCEN). *Health Policy*, 69(3):365-373.

Jones JG, Garrett J, Worthington T. (2004). A videotape series for teaching physicians to evaluate sexually abused children. *J Child Sex Abus*, 13(1):87-97.

Judd F, Cockram A, Davis J, Fahey A, Hodgins G, Jackson H, Scopelliti J. (2003). First year of practice visits for the Rural Depression Anxiety Research and Treatment General Practice program. *Aust J Rural Health*, 11(4):175-180.

Kirkpatrick DL. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco: Berrett-Koehler.

Kjeldmand D, Holmstrom I, Rosenqvist U. (2004). Balint training makes GPs thrive better in their job Patient. *Educ Couns*, 55(2):230-235.

Korn LM, Reichert S, Simon T, Halm EA. (2003). Improving physicians' knowledge of the costs of common medications and willingness to consider costs when prescribing. *J Gen Intern Med*, 18(1):31-37.

Kronz JD, Silberman MA, Allsbrook WC, Epstein JI. (2000). A web-based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by needle biopsy: validation of a new medical education paradigm. *Cancer*; 89(8):1818-1823.

Labelle M, Beaulieu M, Renzi P, Rahme E, Thivierge RL. (2004). Integrating clinical practice guidelines into daily practice: impact of an interactive workshop on drafting of a written action plan for asthma patients. *J Contin Educ Health Prof*, 24(1): 39-49.

Ladouceur R, Goulet F, Gagnon R, Boule R, Girard G, Jacques A, Frenette J, Carrier R. (2003). Breaking bad news: impact of a continuing medical education workshop. *J Palliat Care*, 19(4):238-245.

Lau MA. (2004). Design and feasibility of a new cognitive-behavioral therapy course using a longitudinal interactive format. *Can J Psychiatry*, 49(10):696-700.

Lee E, McNally DL, Zuckerman IH. (2004). Evaluation of a physician-focused educational intervention on medicaid children with asthma. *Ann Pharmacother*, 38(6): 961-966.

Levine HG, Moore DE, Pennington HC. (1984). Continuing education for health professionals: developing, managing and evaluating for maximum impact on patient care. In: Green JS, ed. *Evaluating continuing education and outcomes.* San Francisco, CA: Jossey Bass.

Light RJ, Pillemer DB. (1984). Summing Up the Science of Reviewing Research. Cambridge, Massachusetts and London, England: *Harvard University Press*.

March SK, Bedynek JL Jr, Chizner MA. (2005). Teaching cardiac auscultation: effectiveness of a patient-centered teaching conference on improving cardiac auscultatory skills. *Mayo Clin Proc*, 80(11):1443-1448.

Marguet CG, Young MD, L'Esperance JO, Tan YH, Ekeruo WO, Preminger GM, Albala DM. (2004). Hand assisted laparoscopic training for postgraduate urologists: the role of mentoring. *J Urol*, 172 (1): 286-289.

Marshall JN, Stewart M, Ostbye T. (2001). Small-group CME using e-mail discussions. Can it work? *Can Fam Physician*, 47:557-563.

Marshburn TH, Legome E, Sargsyan A, Li SM, Noble VA, Dulchavsky SA, Sims C, Robinson D. (2004). Goal-directed ultrasound in the detection of long-bone fractures *J Trauma*, 57(2):329-332.

Martling AL, Holm T, Rutqvist LE, Moran BJ, Heald RJ, Cedemark B. (2000). Effect of a surgical training program on outcome of rectal cancer in the County of Stockholm. Stockholm Colorectal Cancer Study Group, Basingstoke Bowel Cancer Research Project. *Lancet*, 356 (9224): 93-96.

McAdam TK, McLaughlin RE, McNicholl B. (2004). Non-touch suturing technique fails to reduce glove puncture rates in an accident and emergency department. *Emerg Med J*, 21(5): 560-561.

Merckaert I, Libert Y, Delvaux N, Marchal S, Boniver J, Etienne AM, Klastersky J, Reynaert C. (2005). Factors that influence physicians' detection of distress in patients with cancer: can a communication skills training program improve physicians' detection? *Cancer*, 104, 411-421.

Mikami Y, Manabe T, Epstein JI, Shiraishi T, Furusato M, Tsuzuki T, Matsuno Y, Sasano H. (2003). Accuracy of gleason grading by practicing pathologists and the impact of education on improving agreement. *Hum Pathol*, 34(7): 658-665.

Montaño, D. E., Phillips, W. R., & Kasprzyk, D. (2000). Explaining Physician Rates of Providing Flexible Sigmoidoscopy. *Cancer Epidemiol Biomarkers Prev*, 9, 665-669.

Mulvey G, Keogh LA, Temple-Smith MJ. (2000). Outcomes of an educational activity with Victorian GPs aimed at improving knowledge and practices in relation to sexually transmissible diseases. *Aust N Z J Public Health*, 24(1):76-78.

Nakajima K, Wasa M, Takiguchi S, Taniguchi E, Soh H, Ohashi S, Okada A. (2003). A modular laparoscopic training program for pediatric surgeons. *JSLS*, 7(1):33-37.

Nayak D, Aronow WS. (2005). Effect of an ongoing educational program on the use of antiplatelet drugs, beta-blockers, angiotensin-converting enzyme inhibitors, and lipid-lowering drugs in patients with coronary artery disease seen in an academic cardiology clinic. *Cardiol Rev*, 13(2):95-97.

Nguyen BH, Nguyen KP, McPhee SJ, Nguyen AT, Tran DQ, Jenkins CN. (2000). Promoting cancer prevention activities among Vietnamese physicians in California. *J Cancer Educ*, 15(2):82-85.

Norton PG, Ginsburg LS, Dunn E, Beckett R, Faulkner D. (2004). Educational interventions to improve practice of nonspecialty physicians who are identified in need by peer review. *J Contin Educ Health Prof,* 24(4): 244-252.

Pagnanelli G, Soyer HP, Argenziano G, Talamini R, Barbati R, Bianchi L, Campione E. (2003). Carboni I. Diagnosis of pigmented skin lesions by dermoscopy: web-based training improves diagnostic performance of non-experts. *Br J Dermatol*, 148(4):698-702.

Pedley D, Brebner E, Rowlands A, Palombo A, Ferguson J. (2003). A role for tele-education in the centralization of accident and emergency services. *J Telemed Telecare*, 9 Suppl 1:S33-34.

Raasch BA, Hays R, Buettner PG, (2000). An educational intervention to improve diagnosis and management of suspicious skin lesions. *J Contin Educ Health Prof*, 20(1): 39-51.

Ray WA, Stein CM, Byrd V, Shorr R, Pichert JW, Gideon P, Arnold K, Brandt KD. (2001). Educational program for physicians to reduce use of non-steroidal anti-inflammatory drugs among community-dwelling elderly persons: a randomized controlled trial. *Med Care*, 39(5):425-435.

Razavi D, Merckaert I, Marchal S, Libert Y, Conradt S, Boniver J, Etienne AM, Fontaine O. (2003). How to optimize physicians' communication skills in cancer care: results of a randomized study assessing the usefulness of posttraining consolidation workshops. *J Clin Oncol*, 15; 21(16): 3141-3149.

Ricordeau P, Durieux P, Weill A, Chatellier G, Vallier N, Bissery A, Fender P, Allemand H. (2003). Effect of a nationwide program of educational outreach visits to improve the processes of care for patients with type 2 diabetes. *Int J Technol Assess Health Care*, 19 (4):705-710.

Rosenblatt MA. (2004). The educational effectiveness of problem-based learning discussions as evaluated by learner-assessed satisfaction and practice change. *J Clin Anesth*, 16(8):596-601.

Sanci LA, Coffey CM, Veit FC, Carr-Gregg M, Patton GC, Day N, Bowes G. (2000). Evaluation of the effectiveness of an educational intervention for general practitioners in adolescent health care: randomized controlled trial. *BMJ*, 320(7229): 224-230.

Sanders MR, Tully LA, Turner KM, Maher C, McAuliffe C. (2003). Training GPs in parent consultation skills. An evaluation of training for the Triple P-Positive Parenting Program. *Aust Fam Physician*, 32(9): 763-768.

Scardovi A, Rucci P, Gask L, Berardi D, Leggieri G, Ceroni GB, Ferrari G. (2003). Improving psychiatric interview skills of established GPs: evaluation of a group training course in Italy. *Fam Pract*, 20(4): 363-369.

Schneider H, Chabikuli1 N, Blaauw D, Funani I, Brugha R. (2005). Improving the quality of STI care by private general practitioners: a South African case study. *Sex Transm Infect*, 81(5):419-420.

Sedlack RE, Petersen BT, Kolars JC. (2005). The impact of a hands-on ERCP workshop on clinical practice. *Gastrointest Endosc*, 61(1):67-71.

Shipman C, Addington-Hall J, Thompson M, Pearce A, Barclay S, Cox I, Maher J, Millar D. (2003). Building bridges in palliative care: evaluating a GP Facilitator program. *Palliat Med*, 17(7):621-627.

Simunovic MI, Paterson CA, Coates AJ, Wright JR, Levine MN. (2004). Pilot study of the quality initiative in rectal cancer strategy. *Dis Colon Rectum*, 47(10):1594-1598.

Sliwa JA, Makoul G, Betts H. (2002). Rehabilitation-specific communication skills training: improving the physician-patient relationship. *Am J Phys Med Rehabil*, 81(2):126-132.

Smits PB, de Buisonje CD, Verbeek JH, van Dijk FJ, Metz JC, ten Cate OJ et al. (2003). Problem-based learning versus lecture-based learning in postgraduate medical education. *Scand J Work Environ Health*, 29(4): 280-287.

Taanila A, Larivaara P, Korpio A, Kalliokoski R. (2002). Evaluation of a family-oriented continuing medical education course for general practitioners. *Med Educ*, 36(3):248-257.

Taylor RS, Reeves BC, Ewings PE, Taylor RJ. (2004). Critical appraisal skills training for health care professionals: a randomized controlled trial. *BMC Med Educ*, 7; 4(1):30.

Thompson C, Kinmonth AL, Stevens L, Peveler RC, Stevens A, Ostler KJ, Pickering RM, Baker NG. (2000). Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomized controlled trial. *Lancet*, 15; 355(9199):185-191.

van Os TW, van den Brink RH, Jenner JA, van der Meer K, Tiemens BG, Ormel J. (2002) Effects on depression pharmacotherapy of a Dutch general practitioner training program. *J Affect Disord*, 71(1-3): 105-111.

Van Zyl DG, Rheeder P. (2004). Physician education program improves quality of diabetes care. *S Afr Med J*, 94(6): 455-459.

Waldorff FB, Almind G, Makela M, Moller S, Waldemar G. (2003). Implementation of a clinical dementia guideline. A controlled study on the effect of a multifaceted strategy. *Scand J Prim Health Care,* 21(3): 142-147.

Watson E, Clements A, Yudkin P, Rose P, Bukach C, Mackay J, Lucassen A, Austoker J. (2001). Evaluation of the impact of two educational interventions on GP management of familial breast/ovarian cancer cases: a cluster randomized controlled trial. *Br J Gen Pract*; 51(471):817-821.

Weir E, Stieb DM, Abelsohn A, Mak M, Kosatsky T. (2004). Design, delivery and evaluation of an email-based continuing professional development course on outdoor air pollution and health. *Med Teach,* 26(2):166-173.

Zick SM, Benn R. Bridging CAM practice and research: teaching CAM practitioners about research methodology. (2004). *Altern Ther Health Med*, 10(3):50-56.