Original Research

A Systematic Review of Evaluation in Formal Continuing Medical Education

JING TIAN, MD; NANCY L. ATKINSON, PHD; BARRY PORTNOY, PHD; ROBERT S. GOLD, PHD, DRPH, FAAHB.

Introduction: *Physicians spend a considerable amount of time in Continuing Medical Education (CME) to maintain their medical licenses. CME evaluation studies vary greatly in evaluation methods, levels of evaluation, and length of follow-up. Standards for CME evaluation are needed to enable comparison among different studies and to detect factors influencing CME evaluation.*

Methods: *A review of the CME evaluation literature was conducted on primary research studies published from January 2000 to January 2006. Studies assessing only satisfaction with CME were excluded, as were studies where fewer than 50% of the participants were practicing physicians. Thirty-two studies were included in the analyses. Determinations were made about evaluation methods, outcome measures, and follow-up assessment.*

Results: *Only 2 of 32 reviewed studies addressed all evaluation levels: physician changes in knowledge and attitudes (level 2), practices (level 3), and improved patient health status (level 4). None of the studies using self-developed instruments (n = 10) provided reliability and validity information. Only 6 studies used validated scales. Twenty studies had a follow-up period of 6 months or less, and 11 had a follow-up period between 1 and 2 years.*

Discussion: *A gold standard for evaluating the effectiveness of CME would include assessment of all 4 levels of evaluation. A valid, reliable, and adaptable CME evaluation questionnaire addressing variables in the second level is needed to allow comparison of effectiveness across CME interventions. A minimum 1-year postintervention follow-up period may also be indicated to investigate the sustainability of intervention outcomes.*

Key Words: *Continuing, education, medical, literature review, evaluation studies, randomized controlled trial*

Introduction

Continuing medical education (CME) seeks to help physicians stay abreast of and accept patient care advances and discontinue less effective strategies.[1] Physicians spend a considerable amount of time in CME to maintain their medical licenses. According to *State Medical Licensure Requirements and Statistics*,[2] 47 of 54 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.[3] Many reviews have been published during the past decade trying to summarize CME evaluation studies to assess their effectiveness.[3–11] Previous reviews of CME evaluations have shown that they vary greatly in evaluation methods, outcome measures, and length of follow-up. These reviews have also found 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.[5]

None of the previous reviews of CME evaluation studies appear to use a framework with which to assess the quality of research methods. The 4-level evaluation model developed by Kirkpatrick in 1994 has been widely used to assess training effectiveness,[12] and Curran and Fleet adapted this model for use in a summative evaluation of CME in 2005 (TABLE 1).[11] According to the adapted model, evaluation should begin with level 1 (participant satisfaction), then as time and budget allow, sequentially assessing levels 2 (knowledge and attitude change), 3 (physician clinical practice change), and 4 (patient outcomes). Each prior level serves as a basis for the next level's evaluation, and each successive level represents a more precise measure of effectiveness and more rigorous, time-consuming analysis.

Dr. Tian: research assistant, Department of Public and Community Health, University of Maryland, College Park; *Dr. Atkinson:* director, Public Health Informatics Research Laboratory, and assistant professor, Department of Public and Community Health, University of Maryland, College Park; *Dr. Portnoy:* Senior Advisor for Disease Prevention, Office of Disease Prevention, Office of the Director, National Institutes of Health, Rockville; *Dr. Gold:* co-director, Public Health Informatics Research Laboratory, professor, Department of Public and Community Health, and dean, College of Health and Human Performance, University of Maryland, College Park.

Correspondence: Jing Tian, MD, Department of Public and Community Health, University of Maryland, Suite 2387 Valley Drive, College Park, MD 20742–2611; e-mail: tianjing@umd.edu.

 DOI: 10.1002/chp.089

Given the wide variaty of methods, measures, and follow-up, an analysis of CME studies using Curran and Fleet's structured framework[11] was conducted to build on the lessons learned and improve CME evaluation study design. The following research questions guided this review:

- What formal CME studies have evaluated changes in physician knowledge and attitudes, clinical practices, and patient outcomes?
- What are the effects of using different randomization strategies on the capacity to measure outcomes?
- What is the reliability and validity of measurement in these studies?
- What follow-up period is recommended to adequately demonstrate CME effectiveness?

Methods

Sample

Randomized controlled trials (RCTs) and well-designed quasi-experimental trials evaluating CME interventions were the focus of this review, building on the review done by Davis and associates on CME evaluation studies published between 1993 and 1999.[3] A MEDLINE search using the MeSH terms *continuing medical education* and *evaluation* was conducted for the period January 2000 to January 2006. Supplementary articles were identified using a search in EBSCOhost, a gateway (interface) to 150 online databases and thousands of e-journals, including Academic Search Elite, Educational Resources Information Center (ERIC), and various health-related databases, with the same key words.[13]

Davis and colleagues' selection criteria were as follows: primary studies; more than 50% of the participants were practicing physicians; RCT of formal CME educational interventions that were didactic, were using interactive educational techniques, or both objective determinations of health professional performance in the practice setting, determinations of health care outcomes, or both.[3] The inclusion criteria for this study were generally consistent with the Davis and associates review, with a few exceptions. This review also included quasi-experimental studies if they compared experimental with control and comparison groups. Only studies that were at or beyond the second evaluation level of the Kirkpatrick's model were included to focus this review on the evaluation methods useful at the upper evaluation levels and beyond measures of physician satisfaction. Last, this review included studies using objective or self-reported data in evaluating health professional performance to help demonstrate the distribution of these 2 strategies and provide information in support of future research. As a result, 32 studies were identified for this review.[15–45]

Analysis

The review entailed a systematic process. First, information was extracted from the original articles and compiled in a review table regarding levels of evaluation, study design, randomization strategy, unit of analysis, length of follow-up, and outcome measures and instrumentation. Second, frequencies and percentages were calculated according to the information in the review table (eg, the levels of evaluation have been summarized according to the modified Kirkpatrick's model). Third, discussion sections for all the studies in the review table were read thoroughly to extract conclusions and recommendations that were applicable to CME evaluation studies in general. They were integrated with the comments that resulted from frequency and percentage tables and presented in the discussion section of this research.

A summary table for all the studies being reviewed was created, with rows representing studies and columns

TABLE 1. Modified Version of Kirkpatrick's Model for Summative Evaluation

Level	Evaluation Focus	Definition
1	Learner satisfaction (reaction)	Evaluates how well participants liked a program using data on participants' perceptions, satisfaction with program objectives, content, instruction, delivery, or instructors
2	Learning outcomes (learning)	Assessment of changes in skills, knowledge, or attitudes among learners, using pretesting and posttesting study designs
3	Performance improvement (behavior)	Information on the extent to which learning has influenced the postlearning behavior or performance of a learner in his or her practice setting; this level attempts to determine whether newly acquired skills, knowledge, or attitudes are being used in the learner's everyday environmentr
4	Patient or health outcomes (results)	Measures tangible results (eg, improving patient health or improving efficiencies) that are influenced by the performance of the learner as a result of participation in the continuing education activity; evaluation at this level is challenging because learners encounter a variety of uncontrollable variables after leaving continuing education programs

Source: From Bloom.[1]

DOI: 10.1002/chp

representing the variables or areas being reviewed. The areas reviewed for the studies (columns) were article name, source, year; participants and study design; length of follow-up; level of evaluations; study outcomes; established instruments; and intervention strategies (see APPENDIX at the end of this article).

Results

Levels of Evaluation

The percentages of studies with different evaluation strategies and levels of evaluation are presented in TABLE 2. Two-thirds of the studies (66%) were RCTs that did not use cluster sampling. Nearly a fifth (19%) were RCTs using clustered sampling techniques; 5 (16%) had quasi-experimental designs with comparison groups but no randomization.

More than half of the studies (n = 21 [66%]) evaluated only 1 level of evaluation, and 9 (28%) evaluated 2 levels (TABLE 2). That most studies focused on only 1 evaluation level might be attributed to investigator interests or to time and resource constraints. However, none of the articles described how the investigators selected the level of evaluation on which they focused.

Only 2 studies (6%) evaluated all 3 levels. Gask and associates' study is 1 of these 2.[23] In this study, physicians were selected at baseline and randomly assigned to experimental and control groups. Five 2-hour sessions were given to the physicians in the experimental group to improve their clinical skills in handling depressed patients. The outcome variables were physicians' recognition of psychological disorders and physicians' attitudes toward depression (level 2); physicians' prescribing patterns and physicians' psychiatry communication skills (level 3); patients' depression status, patients' psychiatric symptoms; and patients' quality of life, patients' satisfaction with consultations, and patients' health service use and costs (level 4).

More than half (21 [66%]) of the selected studies included an evaluation of physician clinical performance, which is a positive finding because the primary purpose of CME is to maintain and improve clinical performance.[46] For example, Gormley and colleagues gave physicians an intervention on shoulder injection techniques.[27] The variables that were evaluated included physicians' confidence in performing shoulder injections, the number of shoulder injections they performed, and the number of shoulder referrals made in the following 6 months.

Randomization Strategy

The review included 32 studies that were randomized either by individuals (66%) or clusters (19%) or nonrandomized groups (16%) (TABLE 3). The variability in study design made comparing and contrasting studies challenging but revealed potential research constraints.

For most studies, physicians were both the unit of randomization (UOR) and the unit of analysis (UOA), and most of these (56%) were RCT studies. These studies randomly assigned physicians to intervention and control groups, and patients were nested within 1 of the experimental groups.

Three studies used patients as the UOR.[14,29,43] In 1 study, for example, patients were assigned randomly to receive general medical care or exposure therapy given by the general practitioners (GPs) after training.[29] Patient data were used to reflect the effectiveness of the exposure therapy and the effectiveness of the GP training for this therapy. However, using patients as the UOR for evaluating physician behavior change may have introduced bias. The treatment offered to control patients could be contaminated by the physicians' experiences of applying the intervention to patients receiving the experimental condition. Thus, the evaluation could underestimate the true effects of the strategies. Hall and colleagues were aware of this potential threat

TABLE 2. Randomized Clinical Trials (RCTs) by Study Characteristics (N = 32)

Study Characteristic	Category	No. (%)
Study design	RCT	21 (66)
	Clustered RCT	6 (19)
	Quasi-experimental	5 (16)
Levels of evaluation	2	4 (12)
	3	10 (31)
	4	7 (22)
	2, 3	5 (16)
	3, 4	4 (12)
	2, 3, 4	2 (6)
Follow-up period, months	Not noted	1 (3)
	≤6	20 (62)
	7–11	0 (0)
	12–24	11 (34)

TABLE 3. Study Design by Unit of Randomization or Comparison (N = 32)

Study Design	Unit of Randomization or Comparison	No. (%)
RCT	GPs	18 (56)
	Patients	3 (9)
Clustered RCT	GPs	2 (6)
	Practices	2 (6)
	Clinics	1 (3)
	Communities	1 (3)
Quasi-experimental	GPs	3 (9)
	Patients	1 (3)
	Hospital	1 (3)

GPs = general practitioners.

DOI: 10.1002/chp

to internal validity and used clinical practice rather than patients as the UOR in their study.[28]

Six studies that were clustered randomized trials used medical practices, clinics, communities, or GPs as the UOR.[15,16,25,26,28,45] Clustered RCTs first randomly assigned practices, GPs, clinics, or communities into intervention and control or comparison groups. Then the participants in different practices, clinics, or communities nested within 1 of the experimental or control groups. For example, 1 study randomized by practice, allocating physicians to intervention or control groups according to a random number list.[28] Studies that randomized on GP first included the patients in the same group to which the GP was assigned. Physician and patient data were then collected from these 2 groups and compared.

Three quasi-experimental studies used GPs, 1 used patients, and another used the hospital as the unit of comparison. There were no significant differences between these quasi-experimental studies and other randomized clinical trials except that different GPs, patients, and hospitals were assigned rather than randomly assigned to groups.

Another issue that emerged in determining randomization strategy was that the UOR was often not consistent with the UOA. For example, 1 study used outcome measures at the practice level to evaluate the interventions aimed at individual GPs.[44] These evaluation results may be misinterpreted because the correct UOA was not used. Analyses focused on different units (physicians, patients, clinics) have different multivariate coefficients of determination (R^2s), and the R^2 will be higher in data that are more aggregated.[47] Therefore, Waldorff and associates' inflated their study's findings by explaining the change in individual GPs' behavior with outcome change measured at the practice level.[44]

Evaluation Instruments

Varied questionnaires, surveys, and scales were used to evaluate outcomes of CME, including physicians' knowledge, beliefs, attitudes, and perceived confidence; patients' satisfaction of consultations and perception of communication skills; and depression level. Half of the 32 studies (n = 16) used questionnaires specific to the clinical domains addressed (TABLE 4). Six (19%) of these 16 studies adapted existing instruments and provided reliability and validity

TABLE 4. Established Instruments Used in CME Evaluation Studies by Content Area

Source, Year	Content Area	Established Instruments
Backhaus et al., 2002[15]	Diagnosis and treatment of chronic insomnia	Structured diagnostic questionnaire
Casebeer et al., 2003[18]	Chlamydia screening	A 21-item knowledge test
Curtis et al., 2000[19]	Back care and therapy skills	Roland-Morris functional disability scale
Delvaux et al., 2005[20]	Cancer communication skills	*Cancer Research Campaign Workshop Evaluation Manual;* 15-item questionnaire
Flores et al., 2002[22]	Diarrhea and cholera case management	Trained observers
Gask et al., 2004[23]	Depression assessment and management	Hamilton Depression score for depression status; psychiatric symptoms general health questionnaire score; quality-of-life SF-36; patient satisfaction with consultations; Visual Analogue Scale (VAS); economic questionnaire
Gielen et al., 2001[25]	Parental injury prevention counseling	Parent exit surveys
Haug et al., 2000[29]	Exposure therapy for social phobia	Mini International Neuropsychiatric Interview; DSM-IV psychiatric diagnoses, CGI-S, Social Phobia Scale; and Marks Fear Questionnaire
Merckaert et al., 2005[32]	Cancer communication skills	Hospital Anxiety and Depression Scale; VAS; *Cancer Research Campaign Workshop Evaluation Manual*
Mulvey et al., 2000[33]	Sexual history taking and STD screening	Baseline and follow-up questionnaire
Razavi et al., 2003[37]	Cancer communication skills	*Cancer Research Campaign Workshop Evaluation Manual* for physicians' communication skills; 14-item survey for patient perceptions of communication skills
Sanci et al., 2000[38]	Adolescent health and patient-provider communication	Questionnaires completed by the general practitioners
Smits et al., 2003[40]	Management of mental health problems	Knowledge tests; performance in practice (self-reports with performance indicators)
Thompson et al., 2000[42]	Depression assessment and management	Hospital anxiety and depression scale for recognition of depression
Waldorff et al., 2003[44]	Clinical practice guideline on dementia identification and diagnostic evaluation	Mailed survey
Watson et al., 2001[45]	Management of familial breast and ovarian cancer cases	A score that was generated by combining responses to 4 questions, poor instrument

DSM-IV = *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition; CGI-S = Clinical Global Impression Scale; STD = sexually transmitted disease.

DOI: 10.1002/chp

information. For example, Merckaert and colleagues' study used the Hospital Anxiety and Depression Scale.[32]

Ten of the remaining studies (31%) developed their own instruments, such as the diffusion and acceptability questionnaire used in Waldorff and associates' study.[44] However, none of the studies using self-developed instruments documented reliability or validity information.

Among the 21 studies that evaluated physicians' clinical performance, 5 (24%) used self-reported behavioral change questionnaires.[19,27,28,38,40] The other 16 studies objectively documented physicians' behavioral change in clinical settings through observation by an external evaluator.[15,22–28,30,32,34,35,37,40,44,45] For example, in Gielen and colleagues' CME intervention in enhancing anticipatory guidance for injury prevention, audiotapes of medical visits and home observations were used in evaluating physicians' counseling skills.[25]

Length of Follow-up

Length of follow-up varied among the studies reviewed (TABLE 2). Twenty studies (62%) had a follow-up period of 6 months or less. Eleven (34%) had a follow-up period of more than 12 months but less than 24 months. One study did not mention the length of follow-up.

Length of follow-up had an impact on the investigators' ability to draw conclusions about effect sustainability. In Watson and associates' study, for example, follow-up evaluation was conducted 3 to 4 weeks after the intervention.[45] However, this short follow-up period was insufficient for assessing whether benefits would persist over time. A shorter follow-up period also did not allow adequate time for integration of knowledge and skills into medical practice or for recognizing changes in patient outcomes.

Discussion

This review confirmed that the evaluation of CME effects beyond physician satisfaction was still not common. Among studies focused on the upper levels of evaluation, some assessed changes in physician medical knowledge and attitudes and physician behaviors, but patient outcomes were least often evaluated. CME studies published between 2000 and 2006 used a variety of questionnaires to evaluate changes in physician knowledge, attitudes, and beliefs specific to the clinical domain they were addressing, but many did not document reliability and validity. Evaluation methods for physician behavior change ranged from self-report questionnaires to objective observations of clinical performance. Medical records were usually used to evaluate changes in patient outcomes. Follow-up periods ranged from immediately after the study to 3 years. The research results pointed to several implications regarding study design of CME evaluation, outcome evaluation methods, and appropriate length of follow-up.

Kirkpatrick's model recommends that evaluation should begin with level 1 and then move sequentially through the levels when possible. Each successive level represents a different measure of the effectiveness of the CME program and increasing impact on clinical significance.[11] Information from each prior level serves as a base for the next level's evaluation. The earlier levels provide the context in which to interpret the results of later levels.[48] Despite the strong rationale for this model, this review found that many studies conducted only 1 level of evaluation. These findings suggest that future research is needed to assess the impact of these interventions throughout each of the levels to demonstrate how immediate outcomes are related to physician behavior and clinical outcomes. Researchers who evaluate isolated levels should provide a rationale for their selections.

In level 2, both attitude and knowledge change need to be evaluated to assess whether the determinants for physician behavior change are in place. These measures can serve as proxy measures for physician behavior change until more rigorous evaluation methods can be implemented. This study found a lack of valid and reliable CME instruments, pointing to the need for psychometric examination and documentation of evaluation instruments in future CME studies. 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.

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. A 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.

In level 3, the high percentage (66%) of studies evaluating physician behavior suggested that many studies have been able to look beyond immediate CME effects to what physicians were doing in their practices. However, many of the reviewed studies used self-reported measures rather than objective measurements to evaluate changes in physician behavior. Objective observations of behavior change in clinical practice should be used instead of self-reported data to prevent recall bias and the halo effect.[16,27] These findings are consistent with those of previous reviews.[4] Future research is needed to establish acceptable measures of physician performance that can be implemented easily.

In level 4, patient health status improvements should be supported by measurable medical indexes. Evaluating all levels of evaluation would further enable investigation of relationships among levels. Future research is needed to assess the effect of CME interventions throughout the levels on patient behavior and clinical outcomes.

RCTs often required complex randomization strategies to obtain input from clinicians and patients, which may have led

DOI: 10.1002/chp

Lessons for Practice

- A gold standard of CME evaluation would include assessment of all levels of evaluation: participant satisfaction; participant knowledge, attitude, and skills supported by a reliable and validated instrument; change in participant performance in the clinical setting supported by objectively observed data; and patient health status improvements.
- A valid, reliable, and adaptable CME evaluation questionnaire addressing variables in the second level (physician knowledge and attitudes) is needed to allow comparison of effectiveness across CME interventions.
- The CME follow-up period should be at least 12 months to detect the intervention effects and to investigate their sustainability.
- Research identifying recommended ways for randomization is needed in future studies.

to biased results and interpretation. The variation across studies also may prevent comparing the effectiveness of different CME programs. Evaluations should avoid using patients as the UOR for evaluating interventions for physician behavior changes to prevent contamination across treatment conditions and underestimating the outcomes.[28] Because different units have different R^2 values, the UOA should match the UOR to avoid misinterpretation or misrepresentation of the results.[47] Further research is needed to recommend appropriate randomization strategies for future evaluations.

Sustainability of the knowledge, confidence, and skills obtained is an issue that must be kept in mind when designing CME evaluations. These findings suggest the need for further study on the optimal follow-up period to assess medical practice changes and sustainability of intervention effects. The evidence here indicated that 12 months may be the minimum time period for follow-up to detect changes in physician behavior. A longer follow-up period and multiple follow-up points are needed to determine the duration of intervention effects, actual transfer of the intervention to clinical practice and patient health outcomes,[19] and strategies needed to prevent the decay of effects.[17,38,42,45] Longitudinal studies would help reveal optimal follow-up periods to determine whether CME content has been translated into practice.

Limitations

Several limitations need to be mentioned when considering the results and implications of this review. First, a biased sample of studies is possible due to publication bias, which increases the likelihood of a false-positive result. Given the restricted inclusion criteria of this review, the studies selected might underrepresent the CME studies published during the same time period. Second, all the studies had a physician participation rate of more than 50%, which was different from most CME studies where a heterogeneous group of clinicians was targeted.

Conclusion

In summary, a gold standard for CME evaluation would include assessment of all 4 levels using reliable and validated instruments and objective measures. Well-reasoned randomization and adequate follow-up periods were also indicated to determine intervention effects and their sustainability.

APPENDIX: Summary Table for CME Evaluation Studies in the Review

Source, Year	Study Design	Participants and Topic	Intervention	Outcomes	Level of Evaluation	Length of Follow-up
Allard et al., 2001[14]	Clustered RCT	99 general practitioners (GPs) and specialists in Canadian practices administering drugs to 266 patients older than 75 yr	Received mailed clinical medication review reports for their patients	Number of potential inappropriate prescriptions given to elderly patients declined but not significantly	3	1 yr
Backhaus et al., 2002[15]	Clustered RCT	16 GPs providing diagnosis and treatment of chronic insomnia to 4,754 patients	1/2-day training at T1; second training after T2, focus on content, little information about strategies	Significant increase in diagnosis rate; more often advised nonpharmacologic treatments and referral to a sleep expert	3	Not reported

(continued)

DOI: 10.1002/chp

APPENDIX: Continued

Source, Year	Study Design	Participants and Topic	Intervention	Outcomes	Level of Evaluation	Length of Follow-up
Bland et al., 2003[16]	RCT	45 physicians managing urinary incontinence (UI) for 668 patients	3-hr CME-accredited course, training in management of UI, patient educational materials, and on-site physician and office support	Rates of assessment and management of existing UI were low in both the control and intervention groups	3, 4	Right after 1-yr intervention
Boakes et al., 2000[17]	RCT	GPs providing palliative care for terminally ill patients	14-wk training program using experiential learning techniques (clinical attachment, mentoring)	Performance-based rather than cognitive education methods had a greater impact on practice outcomes	2	Immediately after training
Casebeer et al. 2003[18]	RCT	180 primary care physicians providing chlamydia screening services	Three 1-hr Web modules; patient brochures, and physician screening guidelines	Significant increase in chlamydia screening knowledge.	2	Immediately after intervention
Curtis et al., 2000[19]	Clustered RCT	31 GPs in primary care practice providing back care and therapy to 295 patients	2-day educational and skill workshops on optimal low back care and manual therapy techniques, 1 mo apart, with a later refresher session (18 hr total)	More patients receiving manual therapy had completely recovered after the first visit and had a more rapid return to functional recovery	4	2, 4, and 8 wk after the index visit
Delvaux et al., 2005[20]	RCT	62 Belgian physicians encouraged to improve communication skills for cancer patients and relatives	Six 3-hr consolidation workshops were conducted after a 2.5-day basic training program	Significant improvement in asking appropriate screening questions, eliciting and clarifying concerns of patients, and in providing less premature information to relatives	3, 4	Immediately after intervention; 5 mo after control
Faber et al., 2005[21]	Quasi-experimental	21 GPs and 20 other health professionals (OHPs) providing low back pain treatment to 56 patients	4-hr training course for GPs and OHPs; physicians learned to work together based on a collaboration protocol	No positive effect on patient outcomes: pain, disability, quality of life, or medical consumption	4	Immediately after training, 3 and 6 mo
Flores et al., 2002[22]	Quasi-experimental	66 course graduates and 66 doctors and nurses providing diarrhea and cholera case management	Mailed information package and provided personal feedback for the questions in the package to the participants	Increased percentage of diarrhea cases assessed correctly and dehydration cases classified; no improvement in rehydration treatment; insignificant improvement in patient counseling	3, 4	2 mo
Gask et al., 2004[23]	Clustered RCT	36 English physicians providing assessment and management of depression for 189 patients	Five 2-hr sessions (lecture, videotape material, role-play, and written materials)	No significant differences in patient depression, general health, or quality of life at 3 mo; at 12 mo, positive effect on quality of life	2, 3, 4	3 and 12 mo following the course
Gerbert et al., 2002[24]	RCT	46 physicians encouraged to improve skin cancer triage skills	Internet-based skin cancer triage intervention	Intervention group scored significantly higher than control group in 9 of 14 outcome measures and maintained for 5 of the 9 outcomes	3	Immediately after intervention

(continued)

DOI: 10.1002/chp

APPENDIX: Continued

Source, Year	Study Design	Participants and Topic	Intervention	Outcomes	Level of Evaluation	Length of Follow-up
Gielen et al., 2001[25]	RCT	31 residents providing injury prevention counseling to 196 families with infants from birth to age 6 mo	1-hr seminar about injury prevention and the American Academy of Pediatrics TIPP materials; 5-hr experiential instruction on injury prevention content and counseling skills	Significant increase in parental injury prevention counseling for 5 of the 6 safety practices and parental satisfaction; no difference in parents' knowledge, beliefs, and home safety behaviors	3, 4	1 year
Goldberg et al., 2001[26]	Clustered RCT	10 communities as the intervention group; spine surgeons, primary care physicians, patients who were surgical candidates, and hospital administrators	30-mo intervention including surgeon study groups, primary care CME conferences, administrative consensus processes, videodiskaided patient decision making, surgical outcomes management	Surgery rates significantly declined in the intervention communities but increased slightly in the control communities	3	Immediately after intervention
Gormley et al., 2003[27]	RCT	40 Irish GPs providing shoulder joint injections	Lecture, small group work, guided practice with patients	Significant increase in confidence in performing shoulder injections and the number performed; no change in the number of shoulder referrals	2, 3	6 mo after both forms of training
Hall et al., 2001[28]	Clustered RCT	19 medical practices encouraged to follow clinical practice guidelines for management of *Helicobacter pylori* eradication	Outreach visit and audit by a pharmacist trained in the techniques of outreach visiting	No significant change in clinicians' prescribing of medicine for *H. pylori*	3	Compared 12-mo period before and after
Haug et al., 2000[29]	RCT	45 GPs providing exposure therapy to 387 patients with generalized social phobia	30 hr in assessing patients with social phobia and conducting exposure therapy training through scoring videotaped interviews on social phobia scales, role-playing	Significant reduction in target complaints to week 12 and week 24, especially among patients receiving a combined treatment of medication (sertraline) and exposure therapy	4	3 and 6 mo
Holroid et al., 2004[30]	Quasi-experimental	Emergency physicians in Canadian province implementing clinical practice guidelines for extremity radiology use treating 6,398 patients	Provincewide dissemination of the Ottawa Ankle Rules; sequential directed education and personalized feedback strategies	No decrease in radiography during the run-in period or sequential directed education and personalized feedback; the use of radiography did not decrease	3	2 yr
Martling et al., 2000[31]	Quasi-experimental	Most surgeons in Stockholm providing patients with colorectal cancer with total mesorectal excision surgery	3 workshops lasting 3 to 4 days consisting of eleven 3- 5-hr video-based live surgery and 2 histopathology sessions with discussions	No differences in 30-day mortality, anastomotic leakage, or complications for patients with curative abdominal resections; significant decrease in local recurrence and cancer-related death	4	2 yr
Merckaert et al., 2005[32]	RCT	58 Belgian French-speaking physicians who are specialists working with cancer patients (part-time or full-time)	1-hr theoretical information course followed by 2 communication skills-training programs: a 2.5-day training and six 3-hr consolidation workshops	Detection of patients' distress associated positively with physicians breaking bad news and using assessment and support skills; no change over time or between groups in ability to assess patient distress	3	Immediately after course; 5 mo after control

(continued)

DOI: 10.1002/chp

APPENDIX: Continued

Source, Year	Study Design	Participants and Topic	Intervention	Outcomes	Level of Evaluation	Length of Follow-up
Mulvey et al., 2000[33]	RCT	444 Victorian GPs providing STD management care	Educational package on STD management	Statistically significant improvements in knowledge and self-reported practice for 4 of the 6 outcomes that were examined	2, 3	2 mo
Nguyen et al., 2000[34]	RCT	20 Vietnamese primary care physicians on cancer screening and prevention knowledge and skills	Multifaceted intervention including cancer screening reminders, CME seminars, paper materials, newsletters, and oncology data-query programs	Performance rates increased significantly for smoking cessation counseling, Pap test screening, and pelvic examinations but not other cancer prevention activities	3	Immediately after intervention
Raasch et al., 2000[35]	RCT	46 Australian family physicians providing assessment and management of suspicious-looking skin lesions	An educational intervention based on audit and feedback with opportunity for reflection on practice	Improvement in providing clinical information on pathology requests, adequate surgical excision of skin lesions, and physician certainty of diagnosis	3	Immediately after intervention
Ray et al., 2001[36]	Clustered RCT	209 physicians providing management of osteoarthritis in 1,566 elderly patients	Face-to-face visit to study physicians by another physician, and reminder placements in the medical records of patients eligible to have NSAID use reevaluated	Modest reduction in days of prescribed NSAIDs use with concomitant increase in acetaminophen use; no significant changes in other study end points	4	1 yr
Razavi et al., 2003[37]	RCT	63 French physician specialists encouraged to improve cancer communication skills	Six 3-hr consolidation workshops conducted after a 2.5-day basic training program	Significant increase in several communication skills: open directive questions, utterances alerting patients to reality, acknowledgments and empathic statements, educated guesses, and negotiations; significant decrease in premature reassurance	2, 3, 4	Immediately after training, 5 mo after the end of basic training for control group
Sanci et al., 2000[38]	RCT	108 self-selected GPs providing primary health care for adolescents	Six weekly 2.5-hr lectures followed by 6 weekly 2-hr case discussion sessions	Significant improvement in all outcomes at 7 mo except rapport and satisfaction ratings; at 13 mo, most were sustained and objective assessment of competence improved; 98% of participants reported a change in practice attributable to the intervention	2, 3	7 and
Sanders et al., 2003[39]	Quasi-experimental	32 GPs providing parental consultations to prevent severe behavioral, emotional, and developmental problems in children	Training in the Triple P-Positive Parenting Program, a behavior-oriented parent consultation skills training program	Greater satisfaction with parent consultation outcomes; significant increase in use of parent consultation skills; significant overall improvement in GPs interaction skills during parent consultations	2	Immediately after training

(continued)

DOI: 10.1002/chp

APPENDIX: Continued

Source, Year	Study Design	Participants and Topic	Intervention	Outcomes	Level of Evaluation	Length of Follow-up
Smits et al., 2003[40]	RCT	118 occupational health physicians providing management of work-related mental health problems	Problem-based learning for intervention group; traditional lecture-based approach for control program	Both programs improved knowledge, but problem-based learning was more effective in improving clinical skills	2, 3	14 mo
Taylor et al., 2004[41]	RCT	145 physicians from South West England developing critical appraisal skills	1/2-day critical appraisal skills training workshop based on the model of problem-based small-group learning	Significant increase in knowledge on appraising evidence and use of Cochrane Library; nonsignificant increases in ability to critically appraise a systematic review article and evidence-seeking behavior	2	6 mo
Thompson et al., 2000[42]	Clustered RCT	169 physicians in 59 practices providing depression screening and management	In-practice education program based on a clinical-practice guideline; videotapes; small-group discussion; role-play	No significant difference in the sensitivity of physicians to depressive symptoms or recovery rates; recognized patients were more likely to improve at 6 wk if they were seen by an educated physician	3, 4	Immediately after intervention
van zyl and Rheeder, 2004[43]	Quasi-experimental	Physicians at 2 tertiary care diabetes clinics in South Africa providing diabetes care and management to 300 patients	Lecture and structured consultation schedule	Significant improvement in quality of care delivered at the clinic as evidenced by fewer clinic visits and longer consultation times for patients	4	1 yr
Waldorff et al., 2003[44]	RCT	727 physicians in 535 local GP practices in Denmark providing diagnostic evaluations for dementia to elderly patients	Seminars, outreach visits, reminders, and CME small-group training	No increase in adherence to guideline recommendations was observed for laboratory tests or cognitive tests to diagnose dementia	3	Immediately after intervention
Watson et al., 2001[45]	Clustered RCT	426 GPs providing genetic risk assessment and referrals for patients with a family history of cancer	In-practice educational session and information package	Significant improvement in correct referral decisions and slight improvement in confidence in the management of individuals with family history of breast or ovarian cancer	2, 3	Immediately after intervention

RCT = randomized controlled trial; TIPP = Training Into Practice Project; STD = sexually transmitted disease; NSAID = nonsteroidal anti-inflammatory drug.

References

1. Bloom BS. Effects of continuing medical education on improving physician clinical care and patient health: A review of systematic reviews. Int J Technol Assess in Health Care. 2005;21:380–385.
2. American Medical Association. State Medical Licensure Requirements and Statistics. Chicago, IL: Data Trace; 2006.
3. Davis DA, Thomson O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education. JAMA. 1999;282(9):867–874.
4. Davis DA, Mazmanian PE, Fordis M, Harrison RV, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. JAMA. 2006;296:1094–1102.

DOI: 10.1002/chp

5. Jaussent S, Labarere J, Boyer JP, Francois P. Psychometric characteristics of questionnaires designed to assess the knowledge, perceptions, and practices of health care professionals with regards to alcoholic patients. Encephale. 2004;30(5):437–446.
6. Wutoh R, Boren SA, Balas EA. eLearning: A review of Internet-based continuing medical education. J Contin Educ Health Prof. 2004;24(1): 20–30.
7. Cauffman JG, Forsyth RA, Clark VA, Foster JP, Martin KJ, Lapsys FX, et al. Randomized controlled trials of continuing medical education: What makes them most effective? J Contin Educ Health Prof. 2002 Fall;22(4):214–221.
8. Thomson O'Brien MA, Freemantle N, Oxman AD, Wolf F, Davis DA, Herrin J. Continuing education meetings and workshops: Effects on professional practice and health care outcomes. Cochrane Database Syst Rev. 2001;(2):CD003030.
9. Hogan DB, Jennett P, Freter S, Bergman H, Chertkow H, Gold S, et al. Recommendations of the Canadian Consensus Conference on Dementia: dissemination, implementation, and evaluation of impact. Can J Neurol Sci. 2001;28(Suppl 1):S115ÆPT2Ø--S121.
10. Amin Z. Theory and practice in continuing medical education. Ann Acad Med Singapore. 2000 Jul;29(4):498–502.
11. Curran VR, Fleet L. A review of evaluation outcomes of Web-based continuing medical education. Med Educ. 2005;39:561–567.
12. Kirkpatrick DL. Evaluating Training Programs: The Four Levels. San Francisco: Berrett-Koehler; 1994.
13. Enyart MG. An interview with Sam Brooks, senior vice-president of sales and marketing for EBSCO publishing. J Business Finance Librarianship. 2004;10(1):27–38.
14. Allard J, Hebert R, Rioux M, Asselin J. Voyer L. Efficacy of a clinical medication review on the number of potentially inappropriate prescriptions prescribed for community-dwelling elderly people. CMAJ. 2001;164(9):1291–1296.
15. Backhaus J, Junghanns K, Mueller-Popkes K, Broocks A, Riemann D, Hajak G, et al. Short-term training increases diagnostic and treatment rate for insomnia in general practice. Eur Arch Psychiatry Clin Neurosci. 2002;252(3):99–104.
16. Bland DR, Dugan E, Cohen SJ, Preisser J, Davis CC, McGann PE, et al. The effects of implementation of the Agency for Health Care Policy and Research urinary incontinence guidelines in primary care practices. J Am Geriatr Soc. 2003;51(7):979–984.
17. Boakes J, Gardner D, Yuen K, Doyle S. General practitioner training in palliative care: An experiential approach. J Palliat Care. 2000;16(2): 11–19.
18. Casebeer LL, Strasser SM, Spettell CM, Wall TC, Weissman N, Ray MN, et al. Designing tailored Web-based instruction to improve practicing physicians' preventive practices. J Med Internet Res. 2003;5(3):e20.
19. Curtis P, Carey TS, Evans P, Rowane MP, Mills GJ, Jackman A. Training primary care physicians to give limited manual therapy for low back pain: Patient outcomes. Spine. 2000;25:2954–2960.
20. Delvaux N, Merckaert I, Marchal S, Libert Y, Conradt S, Boniver J, et al. Physicians' communication with a cancer patient and a relative: A randomized study assessing the efficacy of consolidation workshops. Cancer. 2005;103:2397–2411.
21. Faber E, Bierma-Zeinstra SM, Burdorf A, Nauta AP, Hulshof CT, Overzier PM, et al. 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. 2005;58(1):75–82.
22. Flores R, Robles J, Burkhalter BR. Distance education with tutoring improves diarrhea case management in Guatemala. Int J Qual Health Care. 2002;14(Suppl 1):S47ÆPT2Ø--S56.
23. Gask L, Dowrick C, Dixon C, Sutton C, Perry R, Torgerson D, et al. A pragmatic cluster randomized controlled trial of an educational intervention for GPs in the assessment and management of depression. Psychol Med. 2004;34(1):63–72.
24. Gerbert B, Bronstone A, Maurer T, Berger T, McPhee SJ, Caspers N. The effectiveness of an Internet-based tutorial in improving primary care physicians' skin cancer triage skills. J Cancer Educ. 2002;17(1):7–11.
25. Gielen AC, Wilson ME, McDonald EM, Serwint JR, Andrews JS, Hwang WT, et al. Randomized trial of enhanced anticipatory guidance for injury prevention. Arch Pediatr Adolesc Med. 2001;155(1):42–49.
26. Goldberg HI, Deyo RA, Taylor VM, Cheadle AD, Conrad DA, Loeser JD, et al. Can evidence change the rate of back surgery? A randomized trial of community-based education. Eff Clin Pract. 2001;4(3):95–104.
27. Gormley GJ, Steele WK, Stevenson M, McKane R, Ryans I, Cairns AP, et al. A randomized study of two training programs for general practitioners in the techniques of shoulder injection. Ann Rheum Dis. 2003;62(10):1006–1009.
28. Hall L, Eccles M, Barton R, Steen N, Campbell M. Is untargeted outreach visiting in primary care effective? A pragmatic randomized controlled trial. J Public Health Med. 2001;23(2):109–113.
29. Haug TT, Hellstrom K, Blomhoff S, Humble M, Madsbu HP, Wold JE, et al. The treatment of social phobia in general practice: Is exposure therapy feasible? Fam Pract. 2000;17(2):114–118.
30. Holroyd BR, Wilson D, Rowe BH, Mayes DC, Noseworthy T. Uptake of validated clinical practice guidelines: Experience with implementing the Ottawa Ankle Rules. Am J Emerg Med. 2004;22(3):149–155.
31. Martling AL, Holm T, Rutqvist LE, Moran BJ, Heald RJ, Cedemark B. 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. 2000;356(9224): 93–96.
32. Merckaert I, Libert Y, Delvaux N, Marchal S, Boniver J, Etienne AM, et al. Factors that influence physicians' detection of distress in patients with cancer: Can a communication skills training program improve physicians' detection? Cancer. 2005;104:411–421.
33. Mulvey G, Keogh LA, Temple-Smith MJ. 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. 2000;24(1):76–78.
34. Nguyen BH, Nguyen KP, McPhee SJ, Nguyen AT, Tran DQ, Jenkins CN. Promoting cancer prevention activities among Vietnamese physicians in California. J Cancer Educ. 2000;15(2):82–85.
35. Raasch BA, Hays R, Buettner PG. An educational intervention to improve diagnosis and management of suspicious skin lesions. J Contin Educ Health Prof. 2000;20(1):39–51.
36. Ray WA, Stein CM, Byrd V, Shorr R, Pichert JW, Gideon P, et al. Educational program for physicians to reduce use of non-steroidal anti-inflammatory drugs among community-dwelling elderly persons: A randomized controlled trial. Med Care. 2001;39(5):425–435.
37. Razavi D, Merckaert I, Marchal S, Libert Y, Conradt S, Boniver J, et al. How to optimize physicians' communication skills in cancer care: Results of a randomized study assessing the usefulness of posttraining consolidation workshops. J Clin Oncol. 2003;21(16):3141–3149.
38. Sanci LA, Coffey CM, Veit FC, Carr-Gregg M, Patton GC, Day N, et al. Evaluation of the effectiveness of an educational intervention for general practitioners in adolescent health care: Randomized controlled trial. BMJ. 2000;320(7229):224–230.
39. Sanders MR, Tully LA, Turner KM, Maher C, McAuliffe C. Training GPs in parent consultation skills: An evaluation of training for the Triple P-Positive Parenting Program. Aust Fam Physician. 2003;32(9):763–768.
40. Smits PB, de Buisonje CD, Verbeek JH, van Dijk FJ, Metz JC, ten Cate OJ, et al. Problem-based learning versus lecture-based learning in postgraduate medical education. Scand J Work Environ Health. 2003;29(4):280–287.
41. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: A randomized controlled trial. BMC Med Educ. 2004;4(1):30.
42. Thompson C, Kinmonth AL, Stevens L, Peveler RC, Stevens A, Ostler KJ, et al. Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. Lancet. 2000;355(9199):185–191.

DOI: 10.1002/chp

43. Van Zyl DG, Rheeder P. Physician education program improves quality of diabetes care. S Afr Med J. 2004;94(6):455–459.
44. Waldorff FB, Almind G, Makela M, Moller S, Waldemar G. Implementation of a clinical dementia guideline: A controlled study on the effect of a multifaceted strategy. Scand J Prim Health Care. 2003;21(3): 142–147.
45. Watson E, Clements A, Yudkin P, Rose P, Bukach C, Mackay J, et al. 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. 2001;51(471):817–821.
46. Levine HG, Moore DE, Pennington HC. 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: Jossey-Bass; 1984.
47. Light RJ, Pillemer DB. Summing Up the Science of Reviewing Research. Cambridge, MA, and London, England: Harvard University Press; 1984.
48. Buckley LL, Goering P, Parikh SV, Butterill D, Foo EK. Applying a "stages of change" model to enhance a traditional evaluation of a research transfer course. J Eval Clin Pract. 2003;9(4):385–390.

DOI: 10.1002/chp