Methods and Definition of Terms

Spyridon S. Marinopoulos and Michael H. Baumann

Chest 2009;135;17S-28S
DOI 10.1378/chest.08-2514

CHEST Supplement

CME: ACCP EVIDENCE-BASED EDUCATIONAL GUIDELINES

Methods and Definition of Terms*

Effectiveness of Continuing Medical Education: American College of Chest Physicians Evidence-Based Educational Guidelines

Spyridon S. Marinopoulos, MD, MBA;
and Michael H. Baumann, MD, MS, FCCP

Background: **A core mission of the American College of Chest Physicians (ACCP) is the education of its members, including continuing medical education (CME). The question of what evidence supports the effectiveness of CME activities became central to the ACCP's Educational Resources Division and its education committee.**
Methods: **An application for consideration as a topic for an evidenced-based guideline was submitted to the ACCP Health and Science Policy Committee in 2004. The application was approved contingent on acceptance by the Agency for Healthcare Research and Quality (AHRQ) as a topic for an evidence-based review to be awarded to an AHRQ evidence-based practice center (EPC). The topic was accepted by AHRQ, with a collaborative revision developed by AHRQ and ACCP of the focused questions submitted in the nomination. The AHRQ awarded the evidence review to The Johns Hopkins University EPC (Baltimore, MD). An expert writing panel was assembled comprising methodologists from the EPC, and recommendations were developed from the EPC evidence review and graded using the ACCP system of categorizing the strength of each recommendation and the quality of evidence.**
Conclusions: **This section describes the processes used to develop these guidelines, including identifying, evaluating, and synthesizing the evidence; assessing the strength of evidence; and grading each recommendation.** ***(CHEST 2009; 135:17S–28S)***

Key words: continuing medical education; evidence-based guideline; methods

Abbreviations: ACCP = American College of Chest Physicians; AHRQ = Agency for Healthcare Research and Quality; CME = continuing medical education; EBG = evidence-based guideline; EPC = evidence-based practice center; HSP = Health and Science Policy Committee; KQ = key research question; TEP = technical expert panel

A complete description of the American College of Chest Physicians (ACCP) evidence-based guideline (EBG) development process can be found online[1] and in print.[2] The purpose of ACCP guidelines is to provide members with the tools to practice evidence-based medicine. Guideline development entails collecting all available evidence systematically, summarizing that evidence, and formulating it into useful recommendations.

*From the Johns Hopkins University School of Medicine (Dr. Marinopoulos), Baltimore, MD; and the University of Mississippi Medical Center (Dr. Baumann), Jackson, MS.
Manuscript received October 20, 2008; revision accepted December 10, 2008.

Correspondence to: Spyridon S. Marinopoulos, MD, MBA, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21231; e-mail: smarino1@jhmi.edu
DOI: 10.1378/chest.08-2514

Methods

ACCP Topic Nomination and Acceptance

A core mission of the ACCP is the education of its members, and providing effective continuing medical education (CME) is central to this mission. The ACCP Educational Resources Division identified the need to review the evidence that supports the effectiveness of CME activities and the types of

activities that are most effective in changing physician behavior and, hopefully, patient outcomes. In 2004, the division submitted an application to the ACCP Health and Science Policy Committee (HSP), which oversees the development of ACCP EBGs, to support "the effectiveness of CME" as a topic for an EBG. After approval by the HSP and the ACCP Board of Regents, this topic was nominated in August 2004 to the Agency for Healthcare Research and Quality (AHRQ) for review by an AHRQ evidence-based practice center (EPC).

The topic was accepted by AHRQ in September 2004, with an initial collaborative revision developed by AHRQ and ACCP of the original focused research questions. AHRQ awarded the evidence review to The Johns Hopkins University EPC (Baltimore, MD).

The Johns Hopkins EPC Process

Selection of The Johns Hopkins EPC Team

The Johns Hopkins EPC assembled a research team with broad and complementary backgrounds in methodology, education, and clinical medicine.[3] Members of this team also had extensive direct experience in conducting systematic literature reviews and providing medical education at all levels, including the administration, design, and instruction of CME. The team included the Johns Hopkins associate dean and director of CME (co-principal investigator), the director of The Johns Hopkins EPC, the Johns Hopkins associate dean for curriculum, and directors of CME and non-CME educational activities.

Appointment and Role of Technical Expert and Peer Review Panel

Following the assignment of the topic, AHRQ appointed an independent outside technical expert panel (TEP) in the field of CME whose role was to provide methodological input in shaping the key questions and guidance in the review. A separate peer review group of experts provided further rigorous critical review once the literature synthesis was completed and the report was drafted. Members of the TEP and peer review panel included prominent leaders and experts in the field of CME, including the American Medical Association CME director, the then-current president (2007) of the Society for Academic CME, deans and academics in the field of CME, and two ACCP members (Table 1). TEP input was elicited and provided regularly in the initial planning stages of the review and was particularly valuable in areas where there was lack of consensus regarding methodological choices and where difficult decisions had to be made within the constraints of a fixed budget and strict deadlines.

Table 1— *Technical Experts and Peer Reviewers*

Alejandro Aparicio, MD, Director, Division of Continuing Physician Professional Development, American Medical Association	S. Barry Issenberg, MD, Associate Professor of Medicine; Assistant Dean, Research in Medical Education; Director, Division of Research and Technology; Assistant Director, Center for Research in Medical Education, University of Miami Miller School of Medicine
Michael H. Baumann, MD, MS, Professor of Medicine, Division of Pulmonary and Critical Care Medicine, University of Mississippi Medical Center	Jocelyn Lockyer, PhD, Director, Continuing Medical Education, and Professional Development Associate Professor, Department of Community Health Services, University of Calgary
Frank C. Berry, Continuing Medical Education Director MedChi, The Maryland State Medical Society	Mary Martin Lowe, MA, Director, Education and Improvement, Accreditation Council for Continuing Medical Education
Nancy L. Davis, PhD, Director, Division of Continuing Medical Education, American Academy of Family Physicians	Don Moore, Jr., PhD, Professor of Medical Education and Administration Director, Division of Continuing Medical Education, Vanderbilt University School of Medicine
Robert Galbraith, MD, MBA, Executive Director, Center for Innovation, National Board of Medical Examiners	LTC Lisa K. Moores, MC, USA, Former Chair, Council of NetWorks; Vice Chair, Continuing Education; and Committee Member, Task Force on Performance Measurement, Walter Reed Army Medical Center
James C. Hebert, MD, Chair, Committee on Continuous Professional Development, American College of Surgeons Associate Dean for Graduate Medical Education and Vice Chair for Education, Department of Surgery, University of Vermont College of Medicine	Charles Willis, MBA, Former Director, Department of AMA PRA Standards & Policy Liaison Activities, American Medical Association Administrative Director, Division of Continuing Physician Professional Development

Topic Development: Conceptual Model and Key Questions

In collaboration with AHRQ, ACCP, and the TEP, the EPC developed a conceptual model of the

context, delivery, and outcomes of CME (Fig 1) and identified six key research questions (KQs) to be addressed in the evidence review. Given the constraints of the budget and deadlines for completion of the review, the parties were required to reach consensus about the extent of the initial search and the scope of each key question. Because the initial literature search by the EPC yielded > 60,000 titles, the decision was made to limit the review, using strict exclusion criteria. The EPC was originally requested to complement its search of the primary literature in CME with a separate review of the simulation literature from nonmedical fields (*eg*, aeronautics, space, military) to determine whether additional conclusions applicable to CME could be drawn. This request proved unfeasible because the field is vast, and the decision was made to limit the search on simulation to non-CME medical education only.

Key Questions

The key questions addressed in the review are as follows:

KQ 1. Is there evidence that particular methods of delivering CME are more effective in imparting knowledge to physicians, changing physician attitudes, acquiring skills, changing physician practice behavior, or changing clinical practice outcomes?

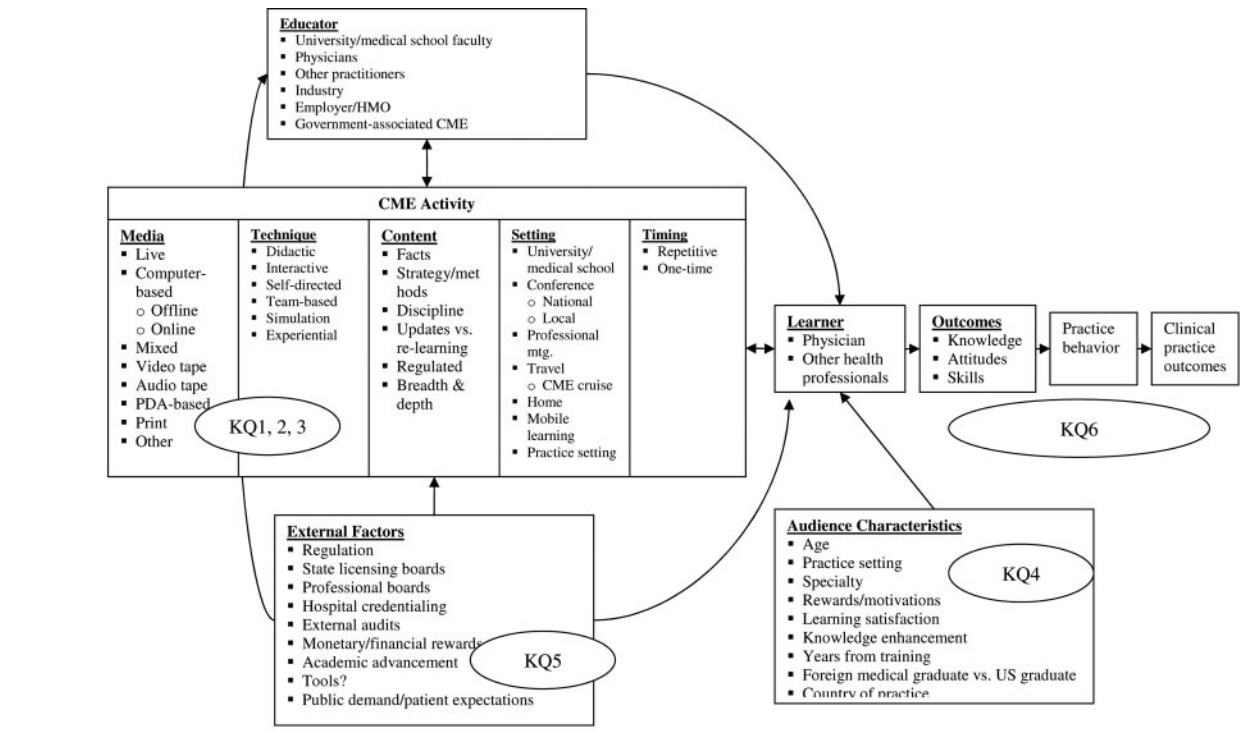
KQ 2. Do changes in knowledge, attitudes, skills, practice behavior, or clinical practice outcomes produced by CME persist over time (≥ 30 days)?

KQ 3. What is the evidence from systematic reviews about the effectiveness of simulation methods in medical education outside of CME?

KQ 4. Which characteristics of the audience by themselves or in combination with other characteristics influence the effectiveness of certain educational techniques?

KQ 5. Which external factors by themselves or in combination with other factors reinforce the effects of CME in changing behavior?

KQ 6. What is the reported validity and reliability of the methods that have been used for measuring the effects of CME in terms of imparting knowledge, changing attitudes, acquiring skills, changing practice behavior, or changing clinical practice outcomes?



Conceptual model of CME constructed by the Johns Hopkins EPC to provide a framework for the Key Questions in conducting their systematic review of the effectiveness of CME.

FIGURE 1. Conceptual model of CME.

To address these key questions, the EPC performed a comprehensive literature search, summarized the available literature, constructed evidence tables, synthesized the evidence, and submitted the report for peer review. In keeping with the directive of the key questions, the EPC focused its literature search, data abstraction, and review on CME aimed at physicians who have completed their training.

Domain Definitions

For the purpose of the literature review and data abstraction, the EPC used the following domain definitions[3]:

- *Knowledge* was defined as any test of physician or CME participant factual knowledge.
- *Attitudes* were any physician or CME participant attitudes; attitudes could include physician attitudes toward a medical topic, comfort level, or satisfaction with the course.
- *Skills* were divided into cognitive (ability to apply knowledge) and psychomotor (*eg*, procedural or physical examination techniques).
- *Practice behavior* referred to any type of physician behavior.
- *Clinical outcomes* were defined as any change in the health status, health-related behavior, or attitudes of *patients* about the physicians for whom the CME intervention was directed.

Search Strategy

The literature search was conducted by an EPC research coordinator experienced in conducting such searches for systematic reviews. These searches require identifying reference sources, formulating a search strategy, and executing the individual searches. To minimize the possibility that important relevant titles would be missed, the search included medical subject heading terms that were relevant to CME from a variety of databases. To minimize the risk of bias in selecting articles for inclusion in the review, a systematic approach for searching the literature with specific eligibility criteria was used.

The search included both electronic and manual searching. Beginning in February 2006, the EPC ran searches of the following databases: MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews, the Cochrane Central Register of Controlled Trials, the Cochrane Database of Abstracts of Reviews of Effects, PsycINFO, and the Educational Resource Information Center.

Two additional databases more specific to the field of CME were considered for reviews of the primary literature: Best Evidence in Medical Education and the Research and Development Resources Base in CME (University of Toronto). The Best Evidence in Medical Education database contained only systematic reviews and not original articles and, therefore, was not searched. Attempts to search the Research and Development Resources Base in CME were unsuccessful and abandoned due to the deadlines of the review process and because searching this database was deemed unlikely to identify additional manuscripts given the thousands already detected.

Manual searching for relevant citations included identifying the 13 journals with the highest number of abstracts and articles included in the review and scanning their tables of contents for relevant citations. To complement the search, reviewers also scanned articles for references of interest that were then compared to the existing database to identify additional citations. This step included scanning the references of articles identified as systematic reviews in CME that were otherwise ineligible for inclusion. Search strategies were specific and tailored to each database.

The literature search for KQ 3 (effectiveness of simulation outside of CME) included a different set of databases that were best tailored to systematic reviews. The search did not include book chapters.

The results of each search were downloaded and imported into research software (ProCite 5; Thomson ResearchSoft; Carlsbad, CA), and duplicate citations were deleted. The articles were uploaded to a Web-based software package developed for systematic review data management (SRS 3.0; TrialStat! Corporation; Ottawa, ON, Canada) that facilitated the subsequent title, abstract, and article reviews and the synthesis of the evidence.

Study Selection

Two independent reviewers conducted title scans in a parallel fashion. The title review phase was designed to capture as many studies as possible. A title was promoted to abstract review if at least one reviewer deemed it eligible. All abstracts were reviewed independently by two investigators. For KQs 1 and 2 (short- and long-term effectiveness of CME) and 4 to 6 (influence of audience characteristics and external factors and evaluation of the tools used to measure CME effectiveness), abstracts were excluded if both investigators agreed that the article met at least one of the following exclusion criteria: not written in English; contained no human data; contained no original data; was a meeting abstract, editorial, commentary, or letter; did not include at least 15 fully trained physicians or fewer than one half of the CME participants completing their training, and results from these physicians were not analyzed separately; did not include training or

education; did not evaluate an educational activity; was published prior to 1981 (the year that Accreditation Council for Continuing Medical Education began accrediting CME) [this date was chosen for feasibility reasons and it was believed that the literature prior to this date was covered in previous reviews and the added value from extending the search prior to 1981 would be limited]; was not conducted in the United States or Canada; did not apply to a key question; did not include data from a concurrent or historical comparison group; or involved quality improvement without an educational activity. Observational studies were included only if they had a comparison group. The search was limited to the United States and Canada to limit the scope of the search and because the EPC team deemed that the differences in medical education and healthcare systems in other countries might make conclusions and recommendations less applicable to the US healthcare system. Manuscripts reporting on at least 15 physicians were chosen because this number would capture the overwhelming majority of studies with a control group to provide adequate statistical power to address the key questions. To qualify for KQ 6 (tools used to measure CME effectiveness), an abstract needed to meet eligibility criteria for at least one other key question.

For KQ 3 (effectiveness of simulation outside of CME), abstracts were excluded if both reviewers agreed that the article met at least one of the following criteria: not written in English; contained no human data; was not a systematic review; was a meeting abstract, editorial, commentary, or letter; did not include medical students or physicians in training; did not include medical training or education; did not evaluate an educational activity; did not involve simulation, virtual reality, manikins, or standardized patients; was published prior to 1990; did not apply to KQ 3; included only fully trained physicians or CME; or did not report separately on the effects of simulation. The cutoff date for inclusion in this review was February 2006.

To capture all possible outcomes across all five domains (knowledge, attitudes, skills, practice behavior, clinical outcomes), no studies were excluded based on the specific methods used to measure such outcomes. Specifically, studies using self-reported outcomes were included in the review because the types of outcomes captured by self-reports often are different from outcomes measured without their use. The validity and reliability of measurements of CME effectiveness was then determined in KQ 6. Articles that passed the abstract review stage underwent an independent parallel review by two investigators who used the same eligibility criteria to determine whether they should be included for full data abstraction and to identify the specific key questions that each article addressed. Articles that passed this stage underwent a full review. Differences of opinion regarding abstract review (24% of instances) and article inclusion or exclusion (10% of articles) were resolved through consensus adjudication.

Data Extraction

Two study investigators reviewed each eligible article and performed full data abstraction and assessment of study quality. A sequential review process was used for data abstraction; the primary reviewer completed standard forms and a second, more experienced reviewer confirmed the completeness and accuracy of these forms. This sequential process was chosen over an independent double review because it saves significant time compared with parallel independent reviews, with minimal errors that would lead to substantial changes in the direction, magnitude, precision, or significance of the pooled estimates for most outcomes.[4] Additionally, the use of an experienced second reviewer and a separate, random audit process further minimized the potential for such errors.

All forms were reviewed by the EPC team during weekly group meetings, piloted using selected articles, and revised to ensure standardization of abstraction between reviewers. To further ensure consistency in data classification and abstraction, an audit process was implemented in which a third experienced reviewer (TD, one of the study's principal investigators) re-reviewed a 10% random sample of articles. Following this process, additional instructions were provided to the reviewers in group session. To facilitate data abstraction, definitions of all terms for media, techniques, and exposures as well as the key questions were available to the reviewers at all times. Issues encountered during the review phase were discussed in detail and resolved by consensus during weekly group meetings. Data were directly abstracted from the article into standardized forms. Each reviewer judged and rated the quality of each study independently. The SRS 3.0 database was used to enter, maintain, and clean the data as well as to create detailed evidence and summary tables.

Data Abstracted To Assess the Effectiveness of CME (KQ 1 and 2)

For all articles containing original data, reviewers extracted information on general study characteristics; CME activity characteristics, including whether the CME activity was accredited; and outcomes. The

details of the general study and CME activity characteristics can be found in the published AHRQ evidence report.[3]

Reviewers were asked to abstract data regarding the main outcome measures for each study, classify them into one or more of the five domains (knowledge, attitudes, skills, behaviors and outcomes), state whether the learning objectives were met, and provide a qualitative summary of the results and the authors' overall conclusions. In the process of reviewing the literature, the EPC found that even if more than one study shared comparable objectives, results such as effect sizes were not reported in a standard fashion. Therefore, it was decided not to perform quantitative metaanalyses of results.

In assessing whether objectives were met, reviewers marked "yes" if half or more of the measures showed improvement, "no" if none of the measures showed improvement, mixed results if only a few of the measures showed improvement or the study reported the results as mixed, no control group if there was not an appropriate control group to answer the question appropriately, and unclear if the results were not clear. If there remained a question about whether the outcome of an objective was mixed, the issue was resolved by agreement at the team level or after a review of the data by the two co-principal investigators.

Data Abstracted From Systematic Reviews on the Effectiveness of Simulation in Medical Education (KQ 3)

Data from systematic review articles were abstracted regarding the types of simulation and comparisons included in the review, types of healthcare professionals included in the review, exclusion criteria, search strategies (types of searches and end date of search), number of articles in the review, outcomes evaluated and the type of objective, meta-analyses conducted, summary of results, subgroup analyses, sensitivity analyses and metaregressions conducted, and overall conclusions. Definitions for the types of simulation were obtained from an internal Johns Hopkins expert (Elizabeth Hunt, MD, Director, The Johns Hopkins Simulation Center).[5]

Data Abstracted To Assess the Influence of Audience Characteristics and External Factors on the Effectiveness of CME (KQ 4 and 5)

A specific form was completed for studies addressing the influence of audience characteristics and external factors that specified the audience characteristic or external factor analyzed, whether a primary goal of the study was to assess the effects of this audience characteristic or external factor, the covariates used in the analysis, and a qualitative summary of the results. Additionally, reviewers abstracted data regarding general study characteristics, CME activity characteristics, outcomes, and study quality.

Data Abstracted To Assess the Validity and Reliability of Methods (KQ 6)

Data regarding the validity and reliability of methods used to assess the effectiveness of CME were abstracted to a specific form. If authors did not identify the specific type of validity or reliability reported, the type was classified based on the definitions from Reed et al.[6] Articles that used a previously validated and reliable method were included if the authors described the method as valid and reliable or described a process or statistic used for psychometric testing. Reviewers also abstracted data regarding general study characteristics, CME activity characteristics, outcomes, and study quality.

Quality Assessment

The quality of articles was assessed differently for clinical trials and systematic reviews. Each original trial underwent a dual, independent review of quality based on the criteria of Jadad et al[7] and included appropriateness of the randomization scheme, appropriateness of the blinding of the assessors, and description of participant withdrawals and dropouts. For each trial, a score between 5 (high quality) and 0 (low quality) was created. Two questions regarding power calculations were added to this form, but the responses to them did not influence the quality score.

In the absence of a standardized system for rating the quality of systematic reviews, The Johns Hopkins EPC assessed these articles using criteria from the Quality of Reporting of Meta-analyses statement[8] as follows: whether the question being addressed by the review was clearly stated, comprehensiveness of search methods used and described in the report, whether inclusion and exclusion criteria were clearly defined and appropriate, whether analyses were conducted to measure variability in efficacy, whether study quality was assessed and done appropriately (using validated instruments), whether differences in how outcomes were reported and analyzed across studies were taken into consideration, whether the study methodology was reproducible; and whether conclusions were supported by the data presented. Additional questions regarding assessment of publication bias were included.

Data Synthesis

Detailed evidence tables containing the information extracted from eligible studies were created for each key question. For KQ 1 and 2 (short- and long-term effectiveness of CME), the results were categorized and sorted based on the media method, the educational technique, and the amount of exposure. Media methods were categorized into single print media (*ie*, the CME activity used only print methods), single live media, single Internet media, other single media, multiple media (*ie*, the CME activity used more than one media method), and single vs multiple media (*ie*, the CME activity for one group used only one media method, and the CME activity for the other group used more than one). Educational techniques were categorized into single technique (the CME activity used only one educational technique), multiple techniques (more than one technique), single vs multiple techniques, and other or not reported. The amount of exposure was categorized into single exposure (*ie*, the CME participants were exposed to the activity on only one occasion), multiple exposures (more than one), single vs multiple exposures, and other or not reported. Investigators used the evidence tables to prepare the text of the report and selected summary tables.

For KQ 6 (methods used to measure the effectiveness of CME), the data were grouped according to similar evaluation methods to facilitate evaluating the validity and reliability of these methods.

Rating the Body of Evidence

At the completion of the review, an evidence grading scheme adapted from the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group[9] was used to grade the quantity, quality, and consistency of the evidence. Grades were given to the bodies of evidence for each of the five domains (*ie*, knowledge, attitudes, skills, practice behaviors, and clinical outcomes). The strength of the study designs was assessed, with randomized controlled trials considered best followed by nonrandomized controlled trials and then observational studies. To assess the quantity of evidence, the EPC team focused on the number of studies with the strongest design. The quality and consistency of the best available evidence was assessed on the following parameters: limitations to individual study quality (using individual quality scores), certainty regarding the directness of the observed effects in studies, precision and strength of findings, and availability (or lack thereof) of data to answer the key question. Evidence bodies (the aggregate of the literature studied for each key question and domain) were classified into the following four basic categories: high grade (further research is very unlikely to change confidence in the estimated effect in the abstracted literature); moderate grade (further research is likely to have an important impact on our confidence in the estimates of effects and *may* change the estimates in the abstracted literature); low grade (further research is very likely to have an important impact on confidence in the estimates of effects and *is likely* to change the estimates in the abstracted literature); and very low grade (any estimate of effect is very uncertain).

Peer Review, Public Commentary, and Publication

A draft of the completed report was sent to the technical experts and peer reviewers as well as to the representatives of AHRQ for critical outside review. In response to the comments of the technical experts, peer reviewers, and AHRQ representatives, revisions were made to the AHRQ evidence report, and a summary of the comments and their disposition was submitted to AHRQ. Once these comments were addressed, the final, peer-reviewed evidence report was published on the AHRQ Web site.[3,10]

Limitations

The AHRQ evidence report[3] provides a comprehensive systematic review of the CME literature, but has several limitations. The major limitations of the report ultimately hinge on the low overall quality of research in CME. In conducting its review, the EPC team encountered many obstacles in its effort to search for and synthesize this body of literature. EPC reviewers encountered a lack of clear definitions of CME, marked heterogeneity of reported outcomes across different audiences and content areas, a lack of valid and reliable tools to measure CME effectiveness, a lack of clear definitions of controls, and a lack of standardized reporting of quantitative data. These limitations of the original literature invariably introduced an additional element of reviewer judgment in conducting the review that was addressed through extensive internal discussion and seeking the input of external experts. However, it quickly became evident that no published consensus existed on many issues, representing a serious deficiency in the state of research in CME that limited the EPC team's ability to draw strong conclusions.

Given the size of the literature, the EPC team, with input from the AHRQ, ACCP, and TEP, made difficult decisions regarding its literature search and methods of data abstraction and synthesis. Different choices based on different expert opinions could have been made with regard to expanding the search strategy; using less stringent exclusion criteria; or

changing the process of abstracting, analyzing, and synthesizing the primary literature. These choices might have, in turn, yielded additional articles in different content and geographic areas, but there is neither published evidence nor expert consensus that such changes would have significantly changed the overall conclusions of the EPC report. Lacking evidence or consensus on whether different methodological choices would have yielded a better review at a reasonable cost, we conclude that further research on this question would be of significant merit. The "Effect on Physician Knowledge" article[11] of these guidelines provides a more detailed and rigorous critique of the strengths and weaknesses of the AHRQ report and outlines the ways in which future systematic reviews in CME might address some of these limitations.

The ACCP EBG Process

Development of the ACCP Guideline

Collaboration With Other Medical Societies and Organizations: Early in the guideline development process, collaboration was sought with other organizations and medical societies who brought important perspectives and complementary expertise. Key organizations represented in the writing panel were the American Medical Association, National Board of Medical Examiners, the University of Toronto Center for Knowledge Translation, and the Accreditation Council for Continuing Medical Education.

ACCP EBG Writing Panel Selection

Under direction and approval from the HSP, a panel of experts in CME was assembled. Several were members of ACCP, and others represented collaborating organizations, including the lead investigators and methodologists of The Johns Hopkins EPC performing the evidence review and creating the AHRQ evidence report. An initial literature review helped to identify content experts having extensive publication experience in the topic area. The selection process identified panel members with the strongest expertise. In addition to clinical and methodological expertise; selection criteria included reliability; ability to work collaboratively; expected substantive contribution to the final product; and gender, minority, and geographic diversity.[2]

Conflict of Interest Management

All writing panel nominees submitted a curriculum vitae or biosketch and a completed standardized conflict of interest form tailored to the topic. All materials were reviewed by the policy and procedures subcommittee of the HSP to determine whether the nominee met the criteria for authorship and had no significant real or perceived conflicts of interest.

The conflict-of-interest policy for ACCP guideline development[1] outlines a process that ensures that disclosed conflicts of interest are properly evaluated and resolved at several key points during the development of the guideline. This policy statement includes an explicit and detailed step-by-step procedure to collect and evaluate the disclosed conflicts respective to the guideline topic and makes recommendations to resolve or manage such conflicts. Guideline panel members were requested to disclose, in writing, conflicts of interest several times during the course of guideline development.[2]

The ACCP does not provide honoraria to members of a guideline writing panel. However, travel expenses for face-to-face meetings related to the guideline development were reimbursed.

EBG Writing Process

Face-to-face meetings of the guideline writing panel were convened in March 2007 and November 2007. At the first meeting, the panel finalized the project's scope and content and received education about the ACCP Grading System[12] and HSP guideline process, including matters of writing format and style. A draft of potential recommendations derived from the AHRQ evidence report[3] was developed prior to the first meeting and served as a template for discussion. Recommendations, including grading, were refined from this draft. Disagreements in recommendations were resolved at this meeting.

Initial drafts of the chapters authored after the March 2007 meeting were circulated so that revisions could be made prior to the November 2007 meeting. The November 2007 meeting occurred after a revised draft of the guideline was completed. This meeting provided wide additional input regarding final recommendations prior to submission to the HSP and education committee. Disagreements in recommendations and chapter discussions were resolved at this meeting.

Grading Recommendations

The writing panel was familiarized with the ACCP Grading System[12] (Table 2) at the outset of the first meeting. The ACCP Grading System was developed by a task force comprising individuals with significant experience in guideline development and grading recommendations. The grading system considers both the quality of the evidence and the balance of benefits to risk and burdens. The ACCP Grading System directly parallels the grading system used by The Johns Hopkins EPC and GRADE.[9] The grading system

Table 2—*Grading Recommendations

Grade of Recommendation and Description	Benefit vs Risk and Burdens	Methodological Quality of Supporting Evidence	Implications
1A. Strong recommendation, high-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	RCTs without important limitations or overwhelming evidence from observational studies	Strong recommendation; can apply to most patients in most circumstances without reservation
1B. Strong recommendation, moderate-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies	Strong recommendation; can apply to most patients in most circumstances without reservation
1C. Strong recommendation, low-quality or very-low-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	Observational studies or case series	Strong recommendation; but may change when higher quality evidence becomes available
2A. Weak recommendation, high-quality evidence	Benefits closely balanced with risks and burdens	RCTs without important limitations or overwhelming evidence from observational studies	Weak recommendation; best action may differ depending on circumstances or patients' or societal values
2B. Weak recommendation, moderate-quality evidence	Benefits closely balanced with risks and burdens	RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies	Weak recommendation; best action may differ depending on circumstances or patients' or societal values
2C. Weak recommendation, low-quality or very-low-quality evidence	Uncertainty in the estimates of benefits, risks, and burdens; benefits, risks, and burdens may be closely balanced	Observational studies or case series	Very weak recommendations; other alternatives may be equally reasonable

*RCT = randomized controlled trial.

combines both low- and very-low-grade evidence found in GRADE into one category. Although designed for clinical recommendations, the ACCP Grading System was adapted for evaluation of CME, despite changing the frame of reference, by regarding benefits vs risk and burden of any recommendation from the patient to the healthcare provider, specifically physicians, or the healthcare system, as appropriate.

EBG Manuscript Review

Based on the proceedings of the November 2007 conference, revisions were made to the guideline and forwarded to several individuals and groups within the ACCP. The HSP reviewed the document for process, consistency, whether the recommendations and grading were appropriate, and content. After the writing panel adequately addressed the critiques provided by HSP and other reviewers, the guideline manuscript was submitted to the ACCP Board of Regents for final approval. Once approved by the ACCP Board of Regents, the manuscript was submitted to *CHEST* for consideration for publication, and an external independent review was performed according to the standard editorial policies of *CHEST*.[2]

Areas Modified From the Original AHRQ Evidence Report

After discussion at the March 2007 guideline panel meeting in Chicago, the panel determined that addi-

Table 3—*The Johns Hopkins EPC Evidence Report: Definitions of Media Methods*

Media Method	Definition
Live	Any CME activity that is conducted in person
Computer-based, offline	Any CME activity that is conducted on the computer but is not conveyed through the Internet (*eg*, CD-ROM)
Internet, real time (*eg*, streaming)	Any CME activity that is conducted in real time through the Internet
Internet, not real time	Any CME activity that is conducted through the Internet but is not conducted in real time
Video	Any CME activity that uses a videotape to convey its message
Audio	Any CME activity that uses an audiotape to convey its message
Handheld	Any CME activity that involves handheld materials (*eg*, laminated card)
Print	Any CME activity that is conducted through educational printed materials or readings

Table 4—*The Johns Hopkins EPC Evidence Report: Definitions of Techniques and Educational Methods*

Technique and Educational Method	Definition
Academic detailing	Detailing provided by an institution or hospital.
Audience response systems	Addresses knowledge objectives. Used in combination with live lectures or discussion groups, these systems are computerized feedback tools that allow the teacher or instructor to pose a question to a large group and receive immediate feedback from each learner, which is collated and presented on a screen. The instructor may choose to alter content based on audience response.
Case-based learning	Addresses higher order knowledge and skill objectives. Actual or authored clinical cases are created to highlight learning objectives; clinical material is presented and followed with questions usually determined by the instructor.
Clinical experiences	Addresses skill, knowledge, and attitudinal objectives. Generally refers to a preceptorship or observership with an expert, as in attending a specialty clinic or an operating room.
Demonstration	Addresses skill and or knowledge (knows how) objectives. Can be presented live or through video or audio media. Teacher determines amount and pace of content.
Discussion group	Addresses knowledge, especially application or higher order knowledge, or affective objectives. Usually requires preparation with readings, or another experience, such as viewing a videotape or a role play. Can be facilitated by instructor, but group often determines content.
Feedback	The provision of information about an individual's performance to learners.
Lecture	Presentation of knowledge content. Live, video, audio, or slide presentation may be available online. Teacher and instructor determines amount and pace of content.
Mentor or preceptor	Addresses higher order cognitive, skill, and affective objectives. Learner is paired with a mentor who may observe, review documentation of performance, advise, coach, and facilitate learning.
Point of care	Addresses knowledge and higher order cognitive objectives (decision-making). Information that is provided at the time of clinical need, integrated into chart or electronic medical record.
Problem-based learning or team-based learning	Addresses higher order knowledge objectives, metacognition, and some skill (group work) objectives. A clinical scenario is presented to a team that identifies the learning objectives, assigns information-seeking tasks, and returns to share information and answer questions about the case. Can be facilitated or nonfacilitated.
Programmed learning	Addresses knowledge objectives. Content is delivered in sequential steps, which are tested with the learner before moving to the next, usually more complicated step. Pace is determined by the learner, but objectives are set by the program (teacher). Can be delivered in text or online.
Readings	Presentation of knowledge content or background for attitudinal objectives. Requires learner to complete; can be done at learner's pace. Teacher or instructor directed or self-directed (*eg*, journals, newsletters, searching online).
Role play	Addresses skill, knowledge, and affective objectives. Learners assume role of patients and clinicians in practicing focused encounters around training problems, usually when standardized patients are unavailable. Encounter may be recorded and reviewed or followed with a discussion group. Rarely used as sole method of education.
Simulation (other than standardized patient or role play)	Addresses knowledge and skill objectives. Ability to simulate potentially addresses higher order integrative objectives, such as responding to an emerging clinical situation, understanding the unfolding of a protein structure, or working in teams. Technology can be used for simulation training of procedures, as in endoscopy virtual reality trainers or anesthesia simulators. Also includes models, such as joint injection and suture. Requires active participation of learner; can use multiple learners in some scenarios.
Standardized patient	Addresses skill and some knowledge and affective objectives. Usually used for communication skills training and assessment, the standardized patient or simulated patient is trained in a specific patient scenario and presentation of a clinical problem. Encounter may be audiotaped or videotaped and timed. Review offers opportunity for reflection and replay of the scenario.
Writing and authoring	Addresses knowledge and affective objectives. Can include authoring test items and participation in test development. Journaling is used frequently for affective objectives and may be followed with discussion groups or review with a mentor.

tional literature should be added to some chapters that were beyond the scope of the AHRQ evidence report.[3] The details of this literature search are outlined in each chapter. In particular, additional citations were included in the simulation chapter because the AHRQ evidence report was limited to a review of reviews.

The guideline writing panel used the AHRQ evidence report to build the recommendations. Some of the titles of the original report were changed to reflect the interests of the guideline writing panel and the ACCP target audience, and chapters were added. Such modifications were made through group

consensus, with the input of the two lead EPC methodologists and investigators, and are noted in their respective sections. The writing panel determined that the changes in terminology better reflect the lexicon of CME. In making these changes, the writing panel was careful to ensure that the new terms accurately reflect the data summarized in the AHRQ evidence report. Specifically, the following changes to the terminology used in the AHRQ evidence report were made:

1. In "Continuing Medical Education Effect on Physician Knowledge Application and Psychomotor Skills"[13] the AHRQ term *skills outcomes* is replaced with *knowledge application*.
2. In "Continuing Medical Education Effect on Practice Performance,"[14] the AHRQ term *practice behavior outcomes* is replaced with *practice performance*, referring to processes of care and not clinical outcomes.
3. The AHRQ evidence report section on physician attitudes was not incorporated in the guideline. Attitudes include areas such as physician satisfaction. The impact of a recommendation regarding a change in an outcome like physician satisfaction may be less meaningful than the impact of recommendations regarding physician knowledge, practice performance, or practice outcomes. Additionally, the EPC experts (Drs. Marinopoulos and Dorman) stated that strong recommendations could not be made about physician attitudes based on the available evidence.
4. No recommendations were made in the guideline regarding KQ 6, which encompasses a methodological question involving reliability and validity of the methods used to measure CME effectiveness. This has important research implications, but limited information precluded the panel from making recommendations.

Definition of Terms

In the AHRQ report,[3] the EPC used definitions for each of the five domains studied (knowledge, attitudes, skills, behaviors, and outcomes). For KQ 1 and 2 (short- and long-term effectiveness of CME) and 4 to 6 (influence of audience characteristics and external factors and evaluation of the tools used to measure CME effectiveness), the EPC used specific definitions for the media (Table 3), techniques and educational methods (Table 4) used in the CME activity. For KQ 3 (effectiveness of simulation outside of CME), the EPC used the specific definitions for simulation (Table 5).

Table 5—*The Johns Hopkins EPC Evidence Report: Definitions of Simulation Types*

Simulation Type	Definition
Full simulation	Whole room or whole patient simulations
Partial task simulation	The use of products to learn or practice a specific skill, such as intubation heads, central venous line chests, intraosseous line legs, or umbilical artery cannulation trainers
Computer simulation	The use of computer programs that allow the student to practice decision-making skills; specific knowledge sets, such as Advanced Cardiac Life Support trainers; and trauma management trainers
Virtual reality	The use of advanced computerized technology to allow students to learn or practice how to perform cardiac catheterizations, colonoscopies, bronchoscopies, ureteroscopies, laparoscopic surgery, hysteroscopy, arthroscopy, ocular surgery, IV line placement
Standardized patient	The use of individuals trained to play the roles of patients, family members, or others to allow students to practice physical examination skills, history-taking skills, communication skills
Role play	Participants play roles of patients, family members, or others to allow practice of communication skills

Conflict of Interest Disclosures

Dr. Marinopoulos was funded by AHRQ contract No. 290-02-0018 to the Johns Hopkins EPC for his role as co-principal investigator of the review of the effectiveness of CME. He also participated in the Genentech Independent Medical Education Advisory Board following the publication of the AHRQ evidence report.

Dr. Baumann has no conflicts of interest to disclose on the subject of this article.

References

1 American College of Chest Physicians Health and Science Policy Committee. HSP and evidence-based medicine. Available at: http://www.chestnet.org/education/hsp/guidelinesProducts.php. Accessed May 1, 2007

2 Baumann MH, Lewis SZ, Gutterman D. ACCP evidence- based guideline development: a successful and transparent approach addressing conflict of interest, funding, and patient-centered recommendations. Chest 2007; 132:1015–1024

3 Marinopoulos SS, Dorman T, Ratanawongsa, et al. Effectiveness of continuing medical education. Baltimore, MD: the Johns Hopkins Evidence-Based Practice Center; evidence report/technology assessment No. 149

4 Buscemi N, Hartling L, Vandermeer B, et al. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol 2006; 59:697–703

5 Hunt EA, Nelson KL, Shilkofski NA. Simulation in medicine: addressing patient safety and improving the interface be-

tween healthcare providers and medical technology. Biomed Instrum Technol 2006; 40:399–404
6 Reed D, Price EG, Windish DM, et al. Challenges in systematic reviews of educational intervention studies. Ann Intern Med 2005; 142:1080–1089
7 Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996; 17:1–12
8 Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement; Quality of Reporting of Meta-analyses. Lancet 1999; 354:1896–1900
9 Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328:1490
10 Agency for Healthcare Research and Quality. Effectiveness of continuing medical education. Available at: http://www.ahrq.gov/downloads/pub/evidence/pdf/cme/cme.pdf. Accessed January 9, 2009
11 Bordage G, Carlin B, Mazmanian PE. Continuing medical education effect on physician knowledge: effectiveness of continuing medical education: American College of Chest Physicians evidence-based educational guidelines. Chest 2009; 135(suppl):29S–36S
12 Guyatt G, Gutterman D, Baumann MH, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force. Chest 2006; 129:174–181
13 O'Neil KM, Addrizzo-Harris DJ. Continuing medical education effect on physician knowledge application and psychomotor skills: American College of Chest Physicians evidence-based educational guidelines. Chest 2009; 135(suppl): 37S–41S
14 Davis D, Galbraith R. Continuing medical education effect on practice performance: American College of Chest Physicians evidence-based educational guidelines. Chest 2009; 135(suppl): 42S–48S