Outcomes Measurement in a National Multimedia Continuing Medical Education Program: Challenges and Opportunities

Jeff C. Huffman, MD, Tim Petersen, PhD, Charissa F. Andreotti, ScB, Lee Baer, PhD, Seth Solway, BA, Robert J. Birnbaum, MD, PhD

Massachusetts General Hospital, Department of Psychiatry, Division of Postgraduate Education, Boston, Massachusetts, USA

ABSTRACT

Background: Continuing medical education (CME) programs face many challenges when they aspire to document improvements in clinician knowledge and behavior. Although the literature is increasing on the measurement of clinically meaningful educational outcomes in CME, such outcomes have not been systematically evaluated in a large, nationwide CME program.

Methods: The Massachusetts General Hospital Psychiatry Academy (MGH-PA) provides CME to a large, diverse nationwide audience. In its 4 years of existence, assessment of learning at MGH-PA events has progressed from simple assessment of isolated facts to more complex evaluations involving assessments of clinically relevant outcomes. We have systematically tracked the development of our outcomes program since the inception of the MGH-PA.

Results: We identified multiple challenges to both the design and practical execution of a plan to measure clinically meaningful educational outcomes in a large-scale nationwide CME program. To address these challenges, we have attempted to apply several innovative approaches to improve the educational rigor of the measures and rates of learner participation. These approaches have led to incremental and substantial improvements in the assessment of clinically meaningful learning among MGH-PA members, although significant barriers to the program’s optimal implementation remain.

Conclusions: Focused approaches to addressing the challenges of educational-outcomes assessment in a large CME program yielded mostly positive results. Additional programmatic adjustments are planned to address ongoing barriers to a full and rigorous evaluation of educational outcomes.

INTRODUCTION

Physicians and other medical clinicians spend the majority of their professional lives as postgraduates, and they use continuing medical education (CME) as a primary method of learning new information [1]. Given the core importance of CME activities in professional development and practice improvement, it is critical that these activities are of high quality, increase knowledge, lead to a change in clinical practice, and improve patient outcomes [2-4].

A key first step in improving CME is to carefully determine what is being learned in the events and to measure what impact such learning has on clinical performance; however, measuring learning outcomes historically has not been a strong point of most CME programs. Often, outcomes are measured through postparticipation questions that evaluate only knowledge about a given topic. Respondents may have had such knowledge prior to attending the lecture, and thus such questions may not measure new learning. There is also typically no follow-up to determine whether the event has led to actual behavioral change in the participants’ practice [5,6].

Recent studies have shown that carefully designed CME programs can be effective in changing clinical performance. In parallel, educational-outcomes measurement in CME has moved from simple inquiries about isolated facts to a more in-depth evaluation of clinical performance and patient outcomes [5,7,8]. Most of these CME activities, however, have occurred at a single site and have involved relatively small numbers of learners, and there have not yet been attempts to measure learning outcomes in large-scale comprehensive CME programs available to a wide audience.

In 2004, we created a broad educational program, the Massachusetts General Hospital Psychiatry Academy (MGH-PA), that provides continuing psychiatric education to a national audience via multiple modalities. The goals of the MGH-PA are (1) to disseminate the best available evidence about psychiatric care to a diverse group of learners and (2) to carefully measure learner outcomes to ensure that the events positively impact clinical decision making.

The overall objective of this article is to present the evolution of educational-outcomes measurement during the 4 years of the MGH-PA’s existence. Specifically, we outline the challenges in measuring clinically meaningful outcomes in a program of this scope, discuss our attempts to respond to these challenges, describe preliminary results from events completed in the first 4 years, and outline future plans for the implementation of more complex and clinically meaningful outcomes measurement.

METHODS

Overview

The MGH-PA was created in 2004 (with events beginning in 2005) as an attempt to expand and improve the educational activities offered through the MGH Department of Psychiatry. From the available evidence about effective CME and adult learning, the design and delivery of our educational offerings have focused on 3 core components: (1) accurate and comprehensive practice-gap analyses, (2) use of sequential and multimodal learning activities,
and (3) use of interactive educational formats [9-12]. The MGH-PA uses a longitudinal curriculum organized along content areas, and this curriculum is divided into 2 learning semesters per year (spring and fall). Each of the major content areas is denoted a module (eg, the Mood Disorders module), and there are typically multiple activities within each module during each semester for creating complementary learning activities in these core areas of psychiatric care. The selection of activities is based on presumed (opinions of subject-matter experts), expressed (participants’ stated needs), and documented (literature-described practice patterns) practice gaps.

During each semester, material is presented via several different educational formats (Table 1). Day-long live symposia consist of several 1-hour lectures (one lecture per content area) and 2 extended discussion panels. These symposia have “traveled” to cities around the country, and an average of 274 attendees participated per event in 2006 (for a total of 6586 attendees for 24 events). Other live events include hour-long Webcasts (roundtable expert discussions of a topic broadcast over the Internet and a satellite network) and expert online forums, which allow learners to directly interact with experts about previously presented material.

The events are also supplemented by a twice-monthly e-mail update to members that provides faculty answers to members’ clinical questions. Written materials and live events are archived on the MGH-PA Web site (http://www.mghcme.org), which allows members access to educational materials at any time. MGH-PA members also interact with one another on the Web site’s Clinical Community Forum to discuss prior events or other relevant topics.

Outcomes Measurement

For a CME program to be truly effective, it must both increase knowledge and improve clinical behavior, and CME programs therefore must carefully evaluate whether these goals are being met. A recent synthesis of the CME literature [3] suggested that a “gold standard” outcomes-measurement program would include evaluation of all levels of the Kirkpatrick and Accreditation Council for Continuing Medical Education (ACCME) frameworks for classifying the outcomes of adult learning [2,13,14] (Figure 1), with a particular emphasis on achieving the highest levels of learning (in practice changes and patient outcomes). In addition, this synthesis recommended systematically assessing the longitudinal, more long-term impact of educational programs on providers’ knowledge and behavior. During the first 4 years of the MGH-PA, we have incorporated the following elements into our outcomes measurement in an attempt to achieve these measurement goals:

• Polling questions. A group of polling questions is generated by the subject-matter experts who created the lectures and by the educational-outcomes group of the MGH-PA. The purpose of these questions is to establish clinicians’ knowledge about existing evidence and current practice patterns related to critical topics in psychiatry (see Table 2 for an example). One polling question is developed for each learning module and administered at the beginning of each lecture. Answers to questions are captured via the use of an electronic audience-response system (ARS) that permits immediate display of results and linkage to anonymous clinical and demographic characteristics of the participants.

• Pre-/postactivity questions. Pre- and postactivity questions evaluate baseline competence and measure intra-activity learning. A pool of multiple-choice questions is generated by faculty experts on the topic and by the external outcomes team (T.P., J.H., R.B., L.B.). Questions focus on clinical principles germane to community-based practice and begin with a brief clinical vignette (see Table 2 for sample questions). One question is ultimately selected per hour-long live lecture, and this question is administered during the event (before the lecture and up to 2 hours after the lecture) via the ARS system.

### Table 1. Description of Core MGH-PA Educational Formats

<table>
<thead>
<tr>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live symposia</td>
<td>Day-long live didactic lectures with hour-long lectures on 6 topics corresponding to 6 longitudinal educational modules. Question-and-answer sessions occur after each lecture, and 2 longer discussion panels at the end of the morning and afternoon sessions discuss additional attendee questions. These events occur in 8 to 12 cities nationwide each semester and are archived on the MGH-PA Web site.</td>
</tr>
<tr>
<td>Satellite broadcasts</td>
<td>Hour-long roundtable live events involving discussion between 3 faculty experts on a given topic. Video clips from other experts are also used to supplement the discussion. The events are accessed via a satellite network and the Internet, and questions from listeners (via phone and the Internet) are fielded by discussants during the event. These events are also archived on the Web site.</td>
</tr>
<tr>
<td>Expert online forums</td>
<td>Live online educational activities that occur approximately 1 month after live symposia and satellite broadcasts. During these events, members participating online send questions about the original lecture/broadcast topic to the lecturer from the original event, who provides direct audio responses and feedback to the participants in real time.</td>
</tr>
</tbody>
</table>
• Follow-up assessment. To assess the degree to which newly learned knowledge is retained, we invite attendees of live symposia to be members of a follow-up cohort. This cohort is contacted via e-mail 6 weeks, 3 months, and 6 months after an event. Follow-up assessment questions represent alternative forms of the pre-/postactivity questions used at the live event, and members respond via an automated survey.

• Assessment of decision-making bias. Decision-making bias and other cognitive errors can adversely influence diagnostic assessments and treatment selection [15,16]. In consultation with heuristics expert Dr. Steven Sloman, we have created surveys of decision-making bias that are administered to participants at each live symposium. These surveys provide case vignettes that attempt to uncover cognitive errors that practitioners may bring to clinical-care encounters. As with our other measures, these results are anonymously linked to relevant provider characteristics to enrich analyses.

RESULTS

During the execution of the outcomes-measurement plan outlined above, we encountered several challenges that required modification of our approaches. Below we highlight critical challenges that arose during the early stages of executing our outcomes plan, along with the modifications we made to respond to these challenges.

Creating Pre-/Postactivity Questions that Assess Broad Clinical Principles rather than Isolated Facts

The field of CME has moved away from a focus on the learning of isolated facts as a primary educational outcome. Instead, cutting-edge CME programs are increasingly focused on measuring change in actual clinical performance [3,5,10,17]. A major challenge faced by the MGH-PA was to accurately assess clinical performance in a manner that was efficient, acceptable to attendees, and feasible in a large, open CME program that involved thousands of participants nationwide. An additional challenge was to educate lecturers about our educational-assessment methods and ensure that the lecturer invariably covered the content of pre-/postactivity questions at each event.

Our initial attempts at pre-/postactivity questions focused on simple facts and were not always directly related to clinical performance (eg, “What was the primary outcome measure of the CATIE trial?”). This approach ensured that the content was covered, but it did not meet the increasing demand for the measurement of clinically related knowledge and behavior.

In response to these initial suboptimal attempts, we then took several steps to improve our question-generation process. This process involved literature review, standardization of questions, and enhanced review of the questions before their use. First, we reviewed the literature on the measurement of clinical performance. Most notably, Peabody and colleagues [18,19] have found that clinical-case vignettes using open-ended responses closely simulated clinical performance, as measured by chart review and the use of simulated patients. In addition, we noted that CME was focused on ensuring that the material learned was rooted in evidence from the literature. In essence, our review revealed that both the form of the question (case-based) and the content (clinically relevant) are exceedingly important.

To ensure that our pre-/postactivity questions more closely simulated Peabody and colleagues’ assessments of clinical performance and to provide consistency across educational-outcomes assessments, we then developed guidelines for each pre-/postactivity question. These guidelines stipulated that each question must be covered in the lecture and have an answer that is evidence-based and that can be directly referenced to the literature. Furthermore, to better assess true clinical performance, we phrased the questions as clinical-case vignettes whenever possible to determine whether learners could extrapolate newly learned information to a patient-care scenario. These changes produced a dramatic increase in educational-outcomes questions being posed as clinical vignettes: from 0% (0/6) in spring 2006 to 83% (15/18) over the next 3 semesters.

In addition to this standardization of questions, we have taken 2 additional steps to improve the quality of questions, by first reviewing them carefully with lectures

<table>
<thead>
<tr>
<th>Model</th>
<th>Stages</th>
<th>Kirkpatrick</th>
<th>ACCME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reaction →</td>
<td>Learning →</td>
<td>Learning/Competence →</td>
</tr>
<tr>
<td>Did the learner attend and enjoy the event?</td>
<td>Did the learner gain new and clinically significant information?</td>
<td>Did the learner apply the newly learned knowledge to patient care?</td>
<td>Did the application of knowledge lead to improved clinical outcomes?</td>
</tr>
<tr>
<td></td>
<td>Behavior →</td>
<td>Performance →</td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Results</td>
<td>Patient Outcomes</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Comparison of the Kirkpatrick and Accreditation Council for Continuing Medical Education (ACCME) educational-outcomes frameworks.
We were therefore able in this large open CME system to implement substantially improved pre-/postactivity questions that had better internal consistency, more rigorous evaluation for question quality, and greater focus on evidence-based facts and case-based clinical performance. Through this process we have been able to assess the rate of “learning” for each module (by comparing the percentages of correct answers to the module’s pre-/postactivity question administered before and after the lecture) and to break down these rates of improvement by discipline (see Figure 2 for an example). We have recognized, however, that despite our efforts to optimize the quality of the questions, the use of a single, multiple-choice question has substantial limitations if one is attempting both to measure the learning of material from an entire lecture and to assess any change in knowledge/clinical performance due to the lecture. If successful, these vignettes should serve as a much richer and clinical performance–matched educational-assessment tool compared with single multiple-choice questions.

### Logistics of Linking Data to Provider Characteristics

Knowledge of the characteristics of the learners is critical for accurately assessing the effectiveness and limitations of an educational initiative. With respect to CME, such information can range from practice characteristics (eg, how frequently a provider sees patients with this condition), educational level and discipline, years of experience, and participation in prior educational events about the topic. A given lecture may be ideal for new practitioners who occasionally see patients with a particular diagnosis, but the same lecture may be too simple for experienced practitioners who have frequently participated in educational activities about the topic. In addition, certain provider characteristics may predispose learners to greater or lesser improvements in performance in response to a certain type of lecture or format.
In a CME program such as the MGH-PA that deals with large numbers of participants and has providers spread throughout the country, there are both practical and ethical issues regarding the linking of educational-outcomes data to individual providers. Practically, it can be difficult to ensure that information obtained from hundreds or thousands of registration forms and the electronic ARS is correctly and efficiently matched and coded by vendor companies and that this information is transmitted correctly to an educational-outcomes team for analysis. Over the last 3 years, we have changed vendors, ARS systems, and statistical programs to improve this complex process, and we have gone from having information that was essentially unusable at the outset of our educational-outcomes measurement process to obtaining all of the desired linked information from an event within 2 weeks of the event.

Confidentiality is another important issue. It is more difficult to ensure confidentiality in a large, open system than in a CME program that is performed in a single setting. As CME becomes increasingly provided in academic medical centers and as provider information becomes increasingly used as part of educational-outcomes research to optimize the quality of CME, the issue of confidentiality and use of personal information is likely to grow in importance. Balanced against this need for confidentiality (and providers' concerns about liability) is the learners' desire to obtain individualized feedback about their knowledge and performance.

After careful consideration of this issue and consultation with legal and research experts, it has become clear that these educational-outcomes assessments—even when used to simply enhance our educational methods—fall under the definition of research. As such, we have obtained approval for our educational-outcomes research (including linking of data with provider characteristics) through our hospital's institutional review board. All personally identifiable information related to educational assessments is made confidential and stored in locked, password-protected files. Provider information is linked to a deidentified “subject” number that disengages the responses from the individual's characteristics. All information is reported in aggregate, and no characteristics or answers of individual providers are reported in any poster, presentation, manuscript, or other scientific communication.

When case information is requested as part of educational activities, such as case conferences, we are careful to deidentify patient information and to change details to mask the case. Participants are informed of this deidentification process when opportunities are given to present case material. Overall, we have had success from a confidentiality standpoint; the next step is to find a way to provide additional individual or group feedback to learners while maintaining this privacy.

**Maximizing On-Site Participation in Pre-/Postactivity Questions**

For many years, CME events often consisted of a lecturer speaking to a passive audience. Current CME programs must reorient both teachers and learners to a process of education that is much more active, requiring more direct effort from attendees during the events and lecturer tolerance of increased interactions with learners. In this context, gaining attendees' full participation in learning assessments has been a substantial challenge for the MGH-PA, given both the intrinsic nature of such assessments and the large scope of the program.

With respect to the pre-/postactivity questions, there have been several barriers to participation. First, the questions are asked at the beginning of the lecture day, and often attendees, who are used to a passive learning model and who have not yet “engaged” with the lecturers, are surprised or put off by having their knowledge assessed.

![Figure 2](image-url)
Similarly, attendees (who are presumably attending because of their perceived knowledge gaps in a given knowledge area) may worry that they will answer incorrectly, especially given that their responses are linked to registration information. Learners are sometimes confused by the fact that postactivity questions are the same as the pretest questions, leading to an “I answered this already” response (especially if the attendee knew the answer prior to the event). They also may wonder why they should bother to answer the question—they may perceive that there will be no direct benefit to them. Finally, the ARS, although relatively straightforward, can at times be confusing to participants, especially at the beginning of the lecture day, before they have had a chance to familiarize themselves with the system.

To address these problems, we have developed a template of introductory remarks at the beginning of live symposia that targets each of these issues. The moderator of the live symposium outlines the MGH-PA model that emphasizes active learning, interactivity with speakers, and learning assessment; he or she explains that this approach may differ from other CME events learners have attended in the past. The moderator emphasizes that participation in the questions helps participants to be more engaged with the lecture and will help us understand how to modify the content or format of lectures to provide an optimal experience for learners. In addition, we have introduced a “warm-up” question for the ARS to ensure that participants can correctly use the system. Finally, we emphasize that lecturers will discuss participants’ answers to the pre-/postactivity questions after they have been completed so that the participants can hear an expert’s opinion about why certain answers are correct or incorrect.

Using this template, we have had fair audience participation in pre-/postactivity questions. In our last semester of events, 50.3% of the participants answered the preactivity question, 60.1% answered the postactivity question, and 37.1% answered both questions. Although this result represents hundreds of responses to each question (and a total of 437 attendees answering both questions), it continues to fall far short of the optimal participation rate. This process has also led to the development of the case conferences noted above. These conferences involve a much more active learning process (including the answering of several questions via ARS during the lecture and several verbal question/answer periods during the lecture) to further engage participants in the assessment and outcomes process.

**Recruitment of an Adequate Sample Size for Follow-up Assessments**

A related, and even greater, challenge for the educational-outcomes arm of the MGH-PA has been to retain subjects in a follow-up cohort to assess longer-term outcomes of educational activities. Participants in the follow-up cohorts are recruited at live events (via their checking off a box on the registration form) and are contacted after the event as outlined in the Methods section. The barriers to recruiting an adequate sample size for this cohort are numerous. CME attendees are typically accustomed to attending an event and having no follow-up from the educational provider, and no additional effort is typically asked of participants after the event. Learners may fear liability/shame if they do not correctly respond to the questions, and the lack of any tangible incentives for participating is another substantial barrier. Finally, contacting participants (who come from many different institutions, rural and urban, often hundreds or thousands of miles from MGH) can be difficult, even with the vast majority of attendees having ready access to e-mail.

To address these barriers, we have emphasized the educational benefits of revisiting the clinically relevant content presented at the initial event and have discussed the ways in which their participation can help to further improve the format and content of future MGH-PA events. We have used small incentives (such as PDF files of current articles) for follow-up cohort participants and have performed surveys to inquire about optimal methods and contact format; however, despite high sign-up rates at the live events (more than 50% of participants typically check the box for the follow-up cohort) and the steps we have taken to address these barriers, our rates of retention in the follow-up cohort have been very low: There were only 43 responses to follow-up questions in the spring 2007 follow-up cohort.

**DISCUSSION**

Over the last 4 years, we have progressively implemented an educational model that has used several key features of adult-learning theory and clinically relevant educational-outcomes measures in a large-scale educational program, the MGH-PA. Slowly and incrementally, we have begun to create outcome measures to assess what attendees have learned, how well they retain the information, and how effectively the new knowledge is applied to clinical situations.

Our progress in this process—which has involved changing the behavior of learners and our curriculum developers alike—has required frequent adjustments at each step. We had to become more knowledgeable about adult-learning theory and others’ success in measuring learning in CME over the course of designing our educational-outcomes measurement program. At the outset of the MGH-PA, our curriculum-development team was not able to immediately implement many of the educationally sound tenets of the educational-outcomes model, and continued revision is clearly needed. In addition, both lecturers and learners have struggled with a model that represents a substantial change in content, format, and rigor of outcomes assessment compared...
with CME programs that have previously been provided in our department and elsewhere. We recognize that CME participants require time to accept the idea that they can be more active in assessing their learning needs and clinical performance.

The vast majority of educational-outcomes strategies employed by the MGH-PA are not new; many have been adapted from reports in the literature that describe novel methods of CME evaluation that appear to be more effective than traditional CME metrics. Our outcomes-measurement methods are based on the Kirkpatrick and ACCME frameworks for classifying learning outcomes [2,13,14] and have been adapted from prior reviews of the CME literature [5,10,11] and a recently published Agency for Healthcare Research and Quality (AHRQ) evidence report on the effectiveness of CME [12].

The MGH-PA adds to this literature in 2 important ways, however. First, to our knowledge this program is the first attempt to formally and systematically use these evidence-based methods of outcomes measurement in a CME program that spans thousands of members, hundreds of events, and multiple formats of teaching. Second, the MGH-PA appears to be the first CME initiative to combine all of these outcomes methods in a single CME program. The majority of reports regarding the enhancement of CME programs describe one or a small number of novel methods, usually provided in a single session or to a small, homogeneous audience [3,20]. The ability of the MGH-PA to apply such methods to a large, diverse audience suggests that it is possible to critically evaluate clinically relevant CME learning outcomes on a large scale.

Despite these gains, there is still much progress to be made with respect to both the educational-outcomes methodology and its execution. Our educational-outcomes efforts with respect to long-term knowledge consolidation and use have also been slowed thus far by limited participation and retention in the follow-up cohorts. Furthermore, despite significant recent gains, an expansion of our educational-outcomes measurement to measure more fully the impact on the very highest levels of learning (ie, direct impact on patient care) would further strengthen our ability to truly ascertain the impact of the CME events on clinical outcomes. Finally, rigorous educational-outcomes measurements will need to be expanded to Webcasts and other educational activities, because they are currently limited to live lecture events.

The generalizability of these educational methods to smaller programs may be somewhat limited, given that smaller CME programs may not have the resources to implement all of these interventions; however, several of the core pieces of the MGH-PA educational-outcomes measurement plan could be used to enhance outcomes measurement in any given program. Furthermore, smaller programs—with fewer participants from a more limited geographic area—may in fact be better equipped to perform certain tasks related to outcomes, such as retention of follow-up cohort members and the use of chart audits to determine the impact on direct patient care.

Finally, another key point is that our outcomes-assessment methods are limited by the attendees’ choice to participate in the assessment process, both during and after the events. Currently, learners at MGH-PA are not required to participate in the ARS questions before and after lectures, nor are they required to participate in the follow-up cohorts. Therefore, these methods may be characterized as evaluation (and retrospective research), rather than as prospective, randomized research trials of optimal rigor. To reach this level of research methodology, we would have to require prospective informed consent of individual learners prior to the activity. Given the scope of our educational events, we have not yet been able to coordinate such an undertaking.

In the future, the MGH-PA educational-outcomes assessment program will aim to address these and other limitations identified by participants, our internal review, and an external audit. Specific changes will include the following:

- The use of case conferences at live events, as described above, that increase learner participation and assess learning with more rigorous open-ended vignettes instead of multiple-choice questions.
- Increased efforts to recruit, retain, and use follow-up cohorts to better study longer-term learning. We are preparing a needs-assessment survey to investigate additional attendees’ reasons for nonparticipation. In addition, we are considering additional incentives to attempt to improve retention. These incentives may include CME credit for participation (pending accreditation approval), raffle for free entry into an educational activity, and individualized learning programs that identify areas of strength and weakness and provide information about additional learning opportunities.
- Expansion of heuristic assessments to assess systematic cognitive errors in clinical care more rigorously. If such errors are prevalent, future MGH-PA activities will focus on this topic.
- Use of video-based teaching and assessment. We have piloted video-based case vignettes as part of live events and plan to study the impact of these video-teaching tools on initial learning and knowledge retention.

In conclusion, CME is a critical source of new information and a nidus for evidence-based practice change for all clinicians. The model of the MGH-PA outlined in this article shows that a large CME program can adapt evidence-based and clinically relevant outcomes-assessment methods within a large multifaceted program and suggests that small and large CME programs may be able to use aspects of this model to improve the quality
of CME activities in other disciplines and settings. The ongoing limitations of the program, in both structure and execution, also show that the delivery of CME is a process that requires persistent attempts at improvement toward the ultimate goal of improving patient outcomes.

ACKNOWLEDGMENTS

The authors thank Dr. Nancy Bennett for her editorial comments on this manuscript.

REFERENCES