BRIEF REPORT

Measuring Continuing Medical Education Outcomes: A Pilot Study of Effect Size of Three CME Interventions at an SGIM Annual Meeting

Saul J. Weiner, MD¹, Jeffrey L. Jackson, MD, MPH², and Sarajane Garten, MA³

¹Department of Medicine and Department of Pediatrics, University of Illinois at Chicago, and the Veterans Affairs Center for Management of Complex Chronic Care, Chicago, IL, USA; ²Uniformed Services University of the Health Sciences, Bethesda, MD, USA; ³Society of General Internal Medicine, Washington, DC, USA.

BACKGROUND: The ACCME is phasing in new criteria for accreditation from 2008 to 2012. These criteria require CME providers to assess the impact of their interventions.

OBJECTIVES: To assess the feasibility of measuring outcomes at a national meeting, the SGIM evaluation committee conducted a pilot assessment of two workshops and one precourse.

DESIGN AND PARTICIPANTS: Session coordinators prepared a five-item questionnaire to assess the knowledge and confidence of participants. The questionnaire was administered pre, immediately post, and 9 months after the educational sessions.

MEASUREMENTS: Changes in performance were calculated as a standardized difference, or effect size.

RESULTS: All three sessions demonstrated initial knowledge acquisition with effect sizes ranging from 0.39 (small) to 0.99 (large) immediately after the sessions. One session demonstrated sustainment of knowledge over the subsequent 9 months while the other two demonstrated decay. Confidence levels decreased following one of the sessions with an effect size of -0.72 (modest effect).

CONCLUSIONS: Effect size measurement of sessions provides quantitative information about their impact on learning and is one way to achieve ACCME compliance. The method, however, poses methodological and logistical challenges that raise questions about the feasibility of tracking learning and retention following a national meeting.

KEY WORDS: CME; accreditation; practice performance; continuum of medical education; quality and improvements in health care.

J Gen Intern Med 24(5):626–9

DOI: 10.1007/s11606-009-0902-3

© Society of General Internal Medicine 2009

BACKGROUND

In September 2006, The Accreditation Council for Continuing Medical Education (ACCME) issued criteria to be phased in from 2008 to 2012, challenging CME providers to employ “assessment or measurement tools...to analyze changes in strategy, performance, or patient outcomes achieved as a result of (their) activities/educational interventions.”^{1,2} We conducted a pilot study to assess the feasibility of surveys measuring the impact of CME provided at the 2006 annual meeting on both short and long term educational outcomes.

METHODS

Selection Criteria

We selected one precourse (4 hours long) and two workshops (90 minutes) using two criteria: 1) include both research and clinical topics, and 2) exclude sessions where evaluations would be used to select junior faculty for awards.

Design Considerations

In planning the evaluation, we considered three issues: preservation of anonymity, ease of administration, and use of available technology. The survey questions were administered to attendees at three time points: (1) shortly after the registration deadline for all who pre-registered (via an email link to a web-based questionnaire), (2) immediately following the CME activity (paper-pencil version of the questionnaire), and (3) nine months following the meeting (again via an email link). Questionnaires were completed anonymously.

Questionnaire Development and Analysis

Session coordinators prepared five multiple choice questions assessing knowledge, skills and attitudes relevant to the goals of the planned activity, and worded so that (a) they could be used both before and after the CME activity to measure change, and (b) one or more questions specifically address practice application. Submitted questions were reviewed by the SGIM Annual Meeting evaluation committee for both clarity and face validity. We limited the questionnaire to 5 questions, because of time considerations. Two of the session coordinators submitted multiple choice questions with one correct answer (out of five options). They also included a question rating attendees confidence: “How comfort-

The opinions or assertions contained herein are the private views of the author and should not be construed as official or as necessarily reflecting the views of the United States Army or the Department of Defense.

Received October 12, 2007

Revised April 16, 2008

Accepted December 23, 2008

Published online March 5, 2009

Table 1. Knowledge and Comfort Scores

		Possible score	Pre-session	Immediately post		9 months	
				All	Subset who took pretest	All	Subset who took pretest
Research workshop	Knowledge score mean (SD)	0–4	2.63 (0.74) (n=9)	2.82 (0.81) (n=17)	2.95 (0.88) (n=9)	3.82 (0.81) (n=17)	3.44 (0.58) (n=9)
	Comfort rating mean (SD)	1–5	3.75 (1.03) (n=9)	2.17 (0.95) (n=17)	3.13 (0.69) (n=9)	3.82 (0.81) (n=17)	3.25 (0.69) (n=9)
Research precourse	Knowledge score mean (SD)	0–4	2.25 (1.3) (n=4)	3.33 (0.51) (n=6)	3.23 (0.68) (n=4)	2.75 (1.5) (n=5)	3.00 (1.41) (n=4)
	Comfort rating mean (SD)	1–5	4.25 (0.5) (n=4)	4.5 (0.83) (n=6)	4.45 (0.75) (n=4)	3.5 (1) (n=5)	4.25 (0.5) (n=4)
Clinical workshop	Knowledge score mean (SD)	0–8	4.92 (1.5) (n=26)	5.23 (1.6) (n=66)	6.06 (1.72) (n=23)	4.92 (1.6) (n=39)	4.95 (1.5) (n=21)

able do you feel planning a study using (this research method)?” with options: “very comfortable,” “somewhat comfortable,” “neither comfortable nor uncomfortable,” “somewhat uncomfortable,” and “very uncomfortable.” The third session coordinator submitted only knowledge questions, and each with more than one potentially correct answer (out of five options); attendees were asked to circle all correct answers. Questionnaires were scored based on the number of items answered correctly.

We calculated three outcomes: knowledge acquisition, knowledge sustainment and changes in comfort. Knowledge acquisition compared scores for the subset of individuals who completed surveys before and immediately after the session. Knowledge sustainment compared scores obtained immediately after the sessions to those 9 months later. Respondents to the 2nd and 3rd rounds of questionnaires indicated whether they had completed previous rounds. This enabled us to distinguish those who completed the pre-meeting questionnaire from those who did not. The changes were calculated as both the absolute change and as a standardized difference, or effect size (ES) (3). Effect sizes are deemed insignificant if <0.2, small if 0.2–0.5, moderate if 0.5–0.8, and large if >0.8.³ These calculations were for the group and not paired analyses. Reliability of responses across the four or five knowledge items, for each session, was assessed using Cronbach’s alpha.

RESULTS

Forty-seven pre-registered participants were contacted of whom 39 (83%) completed the pre-session evaluation online. Eighty-nine attendees of the CME activity completed the surveys (84% response rate) immediately after the session and 61 (59%) completed evaluations 9 months later online. Table 1 presents the average absolute knowledge and comfort scores for each round of questioning. The knowledge items had good internal consistency across the three sessions (Cronbach’s alpha: 0.82, 0.85, and 0.91). Table 2 presents the effect sizes. Although several of the effect sizes indicated the sessions had a positive effect, it is important to note that confidence intervals are wide due to small sample sizes.

Knowledge Acquisition

Participants in all three sessions demonstrated gains in knowledge compared with before the sessions. Participants in the research methods workshop (90-minute session) had a

modest gain (ES: 0.39, 95% CI: –1.34 to 0.59, n=17), the research precourse (8-hour session) had a large gain (ES: 0.99, 95% CI: –0.62 to 2.26, n=5) and the clinical workshop (90-minute session) had a moderate gain (ES: 0.72, 95% CI: 0.12–1.28, n=39).

Knowledge Sustainment

Participants in two of the three sessions had decay in knowledge over the 9 months after the courses: for the clinical workshop, the decay was small (ES: –0.19, 95% CI: –0.59 to 0.20, n=39), and for the research precourse, the decay was moderate (–0.58, 95% CI: –1.70 to 0.71, n=5). Participants in the research workshop had a large gain in knowledge compared to immediately after the session with an effect size of 1.29 (95% CI: 0.48 to 1.94, n=17).

Comfort

Participants in the 8-hour research precourse had a small increase in comfort with the material compared to before the session (ES: 0.32, 95% CI: –1.12 to 1.66, n=5), though that comfort declined over the subsequent 9 months (ES: –1.09, 95% CI: –2.26 to 0.25). In contrast, participants in the 90-minute research workshop reported a decrease in their comfort with the material compared to before the session

Table 2. Effect Size* and 95% Confidence Interval for the Change in Scores on the Knowledge and Comfort Measures

Knowledge		
	Knowledge acquisition	Knowledge sustainment
Research workshop (n=17)	0.39 (–1.34 to 0.59)	1.29 (0.48 to 1.94)
Research pre-course (n=5)	0.99 (–0.62 to 2.26)	–0.58 (–1.70 to 0.71)
Clinical workshop (n=39)	0.71 (0.12 to 1.28)	–0.19 (–0.59 to 0.20)
Comfort		
Time periods	Comfort change Pre-post	Comfort sustainment Post-9 months
Research workshop (n=17)	–0.72 (–1.62 to 0.28)	0.04 (–0.53 to 0.61)
Research pre-course (n=5)	0.32 (–1.12 to 1.66)	–1.09 (–2.26 to 0.25)

*<0.2 No effect, 0.2–0.5 small effect, 0.5–0.8 modest effect, >0.8 large effect

(-0.72, 95% CI: -1.62 to 0.28), with no change in comfort over the next 9 months (ES: 0.04, 95% CI: -0.53 to 0.61).

Logistical Issues

Session coordinators reported that it took "less than an hour," "about an hour" or "several hours," to create the questions. Programming the questions into the meeting's web application management system took 2 hours. Email list development and mailing required only a few minutes. While the online entry of the pre and 9-month post-session surveys facilitated data processing, the pen-pencil version completed at the end of the sessions required manual data entry. Reviewing questions and analyzing data, even for these few workshops took over 3 hours.

DISCUSSION

These cases illustrate a method for assessing the impact of a national meeting session's learning intervention on attendee knowledge and comfort. All three sessions demonstrated knowledge acquisition; two of the three sessions had subsequent decay and one further gain over the subsequent 9 months. One session had an immediate increase in comfort with the material though this comfort declined and one session had a marked decrease in comfort. Attendees in the later session may have realized during the session that the topic was more complex than they realized.

One problem with such evaluations is the inability to assure validity of the measures employed. Our approach relied on the ability and willingness of the session coordinators to generate questions. The questions had modest internal consistency. While these measures were reviewed by the authors and appeared to have face validity (weak evidence), there were no more rigorous validity measures applied. Developing measures with stronger evidence of validity would greatly increase the complexity and expense for meeting organizers. Fortunately, the ACCME at this point is looking only for meetings to attempt to demonstrate some measure of meeting effect and will not likely require such evaluations be rigorous.

Another concern is survey participation rates. The relatively high participation rates in this study may not be generalizable. SGIM members may be unusually motivated to participate since the annual meeting is member driven. Furthermore, we informed participants they were part of an important pilot activity to assess SGIM's options for providing CME. Requiring evaluations as a precondition for CME credit could be one way to enhance participation rates.

Useful lessons came from our experience with the logistics. Although we optimized automation, the available human and technological resources were stretched even for this small pilot and would not be possible for all sessions at the meeting. Logistical issues beyond the traditional meeting evaluation process included: 1) involving session coordinators in questionnaire development, 2) setting up the questionnaires online, 3) obtaining email addresses for the follow up mailing, and 4) data analysis. Expanding this process to include all sessions at a large meeting would require substantial resources. Professional societies are generally run by a small staff and a modest cadre of volunteer members who change from year to year. Our experience suggests that while tracking

short and long term learning outcomes is possible, it may not be feasible.

There are a number of limitations to our analysis. First, while the items had good internal consistency, there was no formal evaluation of the reliability and validity of our measures; hence our data does not allow a rigorous assessment of the impact of the intervention. Second, there is the issue of response bias: response rates varied on the follow up questionnaires; it is possible that those with higher knowledge were more likely to respond. Third, with small samples sizes, the interpretation of effect sizes becomes problematic. The wide confidence intervals indicate that the true effect for some of the sessions is uncertain. Small samples and missing data are analytic problems that are likely to persist with other CME courses.

CONCLUSIONS

It is currently possible to administer a semi-automated process for assessing the short and long term outcomes of a small sample of sessions at the national meeting. This process can uncover useful information about the impact of various educational interventions on attendees. Such an assessment of even just a few sessions, however, requires a significant amount of effort and time for both meeting staff and volunteers. Expanding the method to include a large number of sessions could be overwhelming with current staffing and the available technology.

A central question, then, is whether it is reasonable to expect societies to stay in the CME business given the mounting requirements for providing CME. Recognizing the logistical challenges, ACCME will not be asking CME providers to do pre- and post- test assessments of each session at national meetings. Their requirements are more general, asking providers to demonstrate that they are making an effort to set predefined goals for each meeting, are attempting to assess in some way the extent to which those goals are being achieved, and show they are taking steps to improve subsequent meetings based on this assessment.^{4,5} Nevertheless, given the variability in effect sizes across CME interventions seen even in this small study, further efforts to collect such data — particularly through the refinement of information systems employed at large meetings — could enhance the quality and impact of continuing medical education.

Acknowledgement: This research is supported by the Veterans Administration HSR&D Service, Center for Management of Complex Chronic Care, Center of Excellence, Illinois.

Conflict of Interest Statement: None disclosed.

Corresponding Author: Saul J. Weiner, MD; Department of Medicine and Department of Pediatrics, University of Illinois at Chicago, and the Veterans Affairs Center for Management of Complex Chronic Care, 840 South Wood Street, M/C 718, Chicago, IL 60612, USA (e-mail: sweiner@uic.edu).

REFERENCES

1. See "CME as a Bridge to Quality: Updated Accreditation Criteria," Sept 2006, available at <http://www.cmanet.org/upload/CME%20Expectation%20Criteria.pdf>. Last accessed August 2008

2. **Regnier K, Kopelow M, Lane D, Alden E.** Accreditation for learning and change: quality and improvement as the outcome. *J Contin Educ Health Prof.* 2005;25(3):174–182.
3. **Kazis LE, Anderson JJ, Meenan RF.** Effect sizes for interpreting changes in health status. *Med Care.* 1989;27:S178–S189.
4. **Kopelow M.** Letter from the CEO of ACCME. Accreditation for Learning and Changes. http://www.accme.org/dir_docs/whats_new/aa64ed0e-fbbc-44fd-abad-550a6da7edef_uploadfile.pdf. Last accessed August 2008.
5. **Moore DE.** A framework for outcomes evaluation in the continuing professional development of physicians. In: Davis D, Barnes BE, Fox R, eds. *The Continuing Professional Development of Physicians: From Research to Practice.* Chicago, IL: AMA Press; 2003: 249–74.