Original Research

Evaluation of a Collaborative Program on Smoking Cessation: Translating Outcomes Framework Into Practice

MARIANNA B. SHERSHNEVA, MD, PhD; CHRISTOPHER LARRISON, BA; SHEILA ROBERTSON, MPH; MIKE SPEIGHT, BS

Introduction: *Although evaluating at multiple outcome levels has been proposed for continuing education activities and programs, it is a complex undertaking and is not done routinely, especially in collaborative, multicomponent programs. This article reports on strategies used and results obtained in an evaluation project that examined multiple outcomes of a US-based collaborative, multicomponent smoking cessation educational program for clinicians.*

Methods: *Evaluation was organized conceptually around the 6 levels of an outcomes-based evaluation model and was conducted using registration data, postactivity evaluations, clinical vignettes with questions assessing knowledge and competence in participants and in a comparison group, a commitment to change approach, data from patient charts to assess clinician compliance on 8 performance measures, and tobacco cessation rates. Additional methods included a success case method study of 9 practices participating in performance improvement (PI) activities and assessment of partner collaboration using a written survey and interviews.*

Results: *The program reached more than 43,000 clinicians who participated in a variety of activities. Participants indicated a high level of satisfaction with the program's educational activities and demonstrated higher scores than a comparison group on 6 of 7 competencies. The majority of participants who responded to commitment to change questions reported intended and implemented practice changes consistent with desired outcomes. Performance outcomes of 3 PI activities varied, with greater improvements observed in 1 activity (9.0% to 36.2% improvement across 8 measures). Lower performance outcomes, but a smoking quit rate of 46.8%, was observed in 2 other PI activities.*

Discussion: *The program had an overall positive impact on the measured variables for clinicians and their patients. Use of 1 outcomes assessment framework acceptable to all members of the collaborative, common measures and evaluation techniques, and centralized data repositories contributed to the success of the program evaluation reported here and is recommended to others who are considering a collaborative program evaluation.*

Key Words: *collaborative educational program, evaluation, outcomes assessment, continuing medical education*

Introduction

In 2007, 9 US-based organizations established the Cease Smoking Today (CS2day) partnership for clinician

education, aimed at reducing the number of persons who smoke.[a] The target audience included primary care physicians, cardiologists, pulmonologists, psychiatrists, pharmacists, physician assistants, nurse practitioners, and nurses. The CS2day program was based on a multisource needs assessment, curriculum based on the 5 A's best practice algorithm (ie, Ask, Advise, Assess, Assist, and Arrange),[2] and approaches to continuing education that were known to be effective, such as using active learning strategies[3] and anticipating barriers to change.[4]

The resulting program consisted of 150 live activities, 4 comprehensive performance improvement (PI) projects, 15 enduring activities, 3 educational exhibits, 83 tools for educators and clinicians, and a Web portal

---

[a]The organizations include California Academy of Family Physicians, CME Enterprise, Healthcare Performance Consulting, Interstate Postgraduate Medical Association, Iowa Foundation for Medical Care, Physicians' Institute for Excellence in Medicine, Purdue University School of Pharmacy and Pharmaceutical Sciences, University of Virginia School of Medicine, and University of Wisconsin School of Medicine and Public Health. See the article by Olson and colleagues,[1] "Factors Contributing to Successful Interorganizational Collaboration," in this supplement.

(www.ceasesmoking2day.com). More than 400 education credits were available to learners in 10 credit types, including *AMA PRA Category 1*[TM]; American Academy of Nurse Practitioners (AANP); American Academy of Physician Assistants (AAPA); Accreditation Council for Pharmacy Education (ACPE); and Maintenance of Certification (MOC) Parts II and IV points. The program was funded in part by an unrestricted educational grant from Pfizer.

Evaluating a program of this size and complexity presents interesting challenges. Program evaluation in continuing medical education/continuing professional development (CME/CPD) in the United States has long relied only on enrollment numbers and learner satisfaction.[5] In spite of 20 years of increased emphasis on outcomes and accountability regarding physician performance and clinical improvements,[6,7] only a minority of recent CME/CPD evaluations assessed educational impact at the levels of performance and patient outcomes, and only a small part of this minority assessed impact at all outcome levels.[8] Likely reasons for this situation include high time and resource costs[8]; multiple influences on outcomes, which limits attribution[9]; insufficient assessment procedures[6,10]; and difficulty obtaining access to data.[11]

Members of the CS2day collaborative wanted the program to impact the significant public health problem of smoking beyond attendance and satisfaction measures. It is increasingly recognized in the CME/CPD field that establishing collaborative arrangements among CME/CPD providers and agencies specialized in educational technologies[12] allows the resulting collaboration to leverage, combine, and capitalize on each other's strengths.[13] Collaboration constitutes a promising way to reduce some of the barriers to comprehensive evaluation through the synergy that would result from complementary capabilities and resources. Because the CS2day program was being offered through the efforts of a collaborative, it seemed to us that there would be resources and expertise available within the collaborative that would address some, if not all, of the reasons why the entire hierarchy of outcomes using multiple indicators is not used more regularly.

Thus, we undertook to evaluate a multicomponent educational program at multiple levels of desired outcomes using the resources and expertise available among the collaborative partners. This article describes our efforts, reporting the strategies that we used and the results that we found.

Methods

The major purpose of summative evaluation of the CS2day program was to document changes in clinicians' competence and performance with respect to the 5 A's best practice algorithm (ie, Ask, Advise, Assess, Assist, and Arrange),[2] and the changes in the quit rates of patients who smoked. Other purposes were to collect data for comparisons among various program components and learn how chosen evaluation approaches worked when applied to a complex collaborative

program. Formative evaluation was continuous and aimed at improvement of ongoing activities and partner collaboration. All findings were to be used in decision making about future programming.

Both formative and summative evaluations were conceptually organized around the 6 levels of the outcomes-based CME evaluation model proposed by Moore in 2003[14] (see Table 13-1). Similar to other multilevel models, such as Kirkpatrick's 4 levels of training evaluation,[15] Moore's model reflects logical progression from learning to learner performance, and to outcomes of their performance, but it specifies "patient health status" and "population health status" as the 2 highest levels, making it most applicable to the health care domain.[b] To conduct activity evaluations, we used evaluation instruments that we developed collaboratively by adopting and adapting existing forms or creating new ones.

To document outcomes at Level 1—Participation,[14] the partners tracked attendance and participant-specific information, including degree and specialty, in a common database. Level 2—Satisfaction[14] was assessed through standard postactivity questions about perceived appropriateness of content, faculty effectiveness, and the effect on clinical practice.

Level 3—Learning[14] refers to changes in knowledge, skills, attitude, and competence, where *competence* is defined as the capability to act.[5] Outcomes at this level were assessed using clinical vignettes.[16] Fifteen developed vignettes were categorized by competency derived from the 5 A's algorithm—Ask about tobacco use, Advise to quit, Assess readiness to quit, Assist with cognitive/behavioral strategies, Assist with medication, Assist with relapse prevention, and Arrange for follow-up. Each vignette had 2 to 4 questions focused on knowledge or competence. The selection of vignettes for each activity was based on the target audience and learning objectives. A comparison group of 157 physicians who did not participate in CS2day activities answered the vignettes applicable to their specialties. When analyzing data, responses to vignettes were scored, the mean score for each vignette was calculated, mean scores were grouped by competency, and the weighted mean score was calculated for each competency based on the number of responses. For each competency, the difference between participants and the comparison group was documented as the percentage of participant mean score exceeding/being less than the comparison mean score.

Outcomes at Level 4—Performance[14] were assessed using 2 methods: a commitment-to-change (CTC) approach and a registry to track performance data on measures. We used the CTC approach with and without follow-up[17] as a proven surrogate measure of actual performance change.[18,19] In some activities, we embedded a request to list intended changes resulting from attending the activity into the postactivity

---

[b]The 2003 outcomes framework[14] was expanded by Moore and colleagues in 2009[7] but the CS2day partners made a decision to stay with the 2003 framework that was consistent with the original evaluation plan.

evaluation. In a smaller number of activities, we complemented this approach by (1) collecting intended changes post activity from a sample of participants using a separate CTC card; (2) reminding them 4 to 6 weeks post activity of their commitments; (3) asking them if they were able to accomplish their commitments and, if not, what barriers prevented them from accomplishing their commitments; and (4) interviewing some of them about their commitments and barriers they might have encountered. In one PI activity, participants were asked about changes in practice only at follow-up 3 months post workshop.

The second method of assessing performance outcomes was a patient registry. Eight CS2day clinical performance measures derived from the tobacco cessation guideline[2] were used by participants of PI activities to assess their performance by entering data from charts into either the CareMeasures™ registry or the CME 360® registry.[c]

The CareMeasures™ registry also allowed for tracking of changes in patient smoking status (eg, when a smoker turned into a former smoker). Thus, patient health status data (Level 5—Patient Health[14]) and, cumulatively, population health status data (Level 6—Population Health[14]) were available.

All certified activities were to be evaluated up to Level 4—Performance[14] and PI activities at all 6 levels. Quantitative outcomes data were to be analyzed by activity and, where possible, aggregated to see the overall impact of the CS2day activities. Statistical analysis was limited to descriptive statistics.

Two additional evaluation components included a success case method (SCM) study and formative assessment of partner collaboration. The SCM study explored the mechanisms linking education and practice changes in the PI activities.[20] Collaboration assessment included a written partner survey[21] and interviews about collaboration experience conducted by an external evaluator.

CS2day evaluation was a shared responsibility. Data from activities sponsored by the partners were collected into centralized databases and analyzed by a multiorganizational team. At the end of the second year of the initiative, this team prepared a detailed final report and several supplemental reports[d] for use by the partners, program planners, faculty, and the grantor.

Results

Outcomes were tracked consistently in every activity at multiple evaluation levels (TABLE 1). The majority of the following results were taken from the final report, which contained data as of January 2010. However, outcomes of the PI activities and TABLE 1 reflect updated data through March 2011.

---

[c]The CareMeasures™ registry (www.encompassonline.com) is supported by one partner (Iowa Foundation for Medical Care). The CME 360® registry is supported by CECity (www.cecity.com).
[d]Copies of the final report/supplemental reports are available upon request.

Level 1—Participation

The CS2day initiative reached more than 43,000 clinicians from all 50 states via certified educational activities, educational exhibits, and tool dissemination through March 2011. Based on data from 13,971 participants who participated in certified activities through January 2010 and responded to a question asking about their profession, 64% were physicians (the majority of whom [80%] represented the main target specialties), 14% were nurse practitioners and nurses, 10% were pharmacists, 8% were physician assistants, and the remaining 4% were other health care professionals.

Level 2—Satisfaction

A pool of satisfaction questions was available from which activity planners could select questions for individual activity evaluations. Overall, participants indicated that they were satisfied with the program's educational activities. For example, participants seemed to agree that the CS2day activities would make them more effective in clinical practice, with a mean score of 4.1 on a scale from 1 = low to 5 = high (5,472 responses). Also, participants reported that the curriculum was appropriate to their clinical practice: 4.2 on a scale from 1 = low to 5 = high (7,948 responses).

Level 3—Learning

Outcomes at Level 3—Learning[14] were drawn from evaluations where clinical vignettes were used (FIGURE 1). Six of the 7 competencies showed participant scores exceeding the comparison group scores.

Level 4—Performance

Many participants who responded to CTC questions reported intended and implemented practice changes that were consistent with desired outcomes. Because different CTC instruments were used across activities, it was not possible to aggregate all CTC data into 1 analysis. We provide several example results below where data were aggregated for activities using the same instrument.

The CTC approach was integrated into state chapter meetings of the American Osteopathic Association, a symposium sponsored by the American Academy of Nurse Practitioners, and a symposium sponsored by American Academy of Physician Assistants. Post activity, 1,071 of 2,829 participants (38%) of these events completed a CTC form, and 576 (54%) of the respondents state chapter meetings indicated a total of 813 intended changes. These changes were in line with the educational objectives and competencies, with the largest proportion of change statements related to advising patients to quit (28%) and assisting patients in developing a quit plan (28%). At follow-up, 215 participants who responded confirmed that 227 (82%) of 277 changes that they initially planned were implemented (FIGURE 2).

TABLE 1. CS2day Program Evaluation by Level

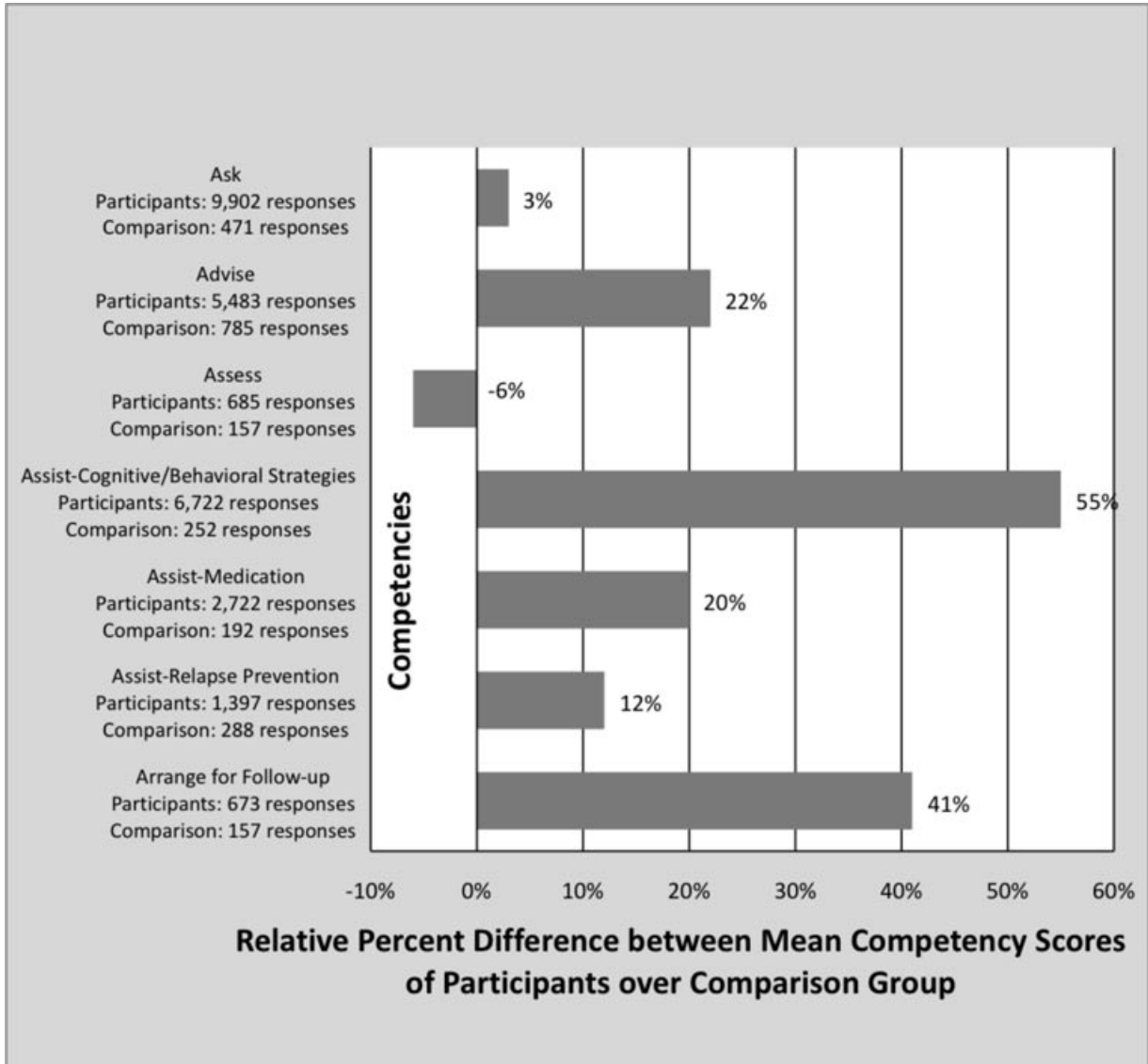| Evaluation Level	Level 1 Participation	Level 2 Satisfaction	Level 3 Learning	Level 4 Performance			Level 5 Patient Health	Level 6 Population Health	Others
Activity/Series of Activities (# of Activities) or Event/Resource	Participant Number	Post Activity Evaluation Questions	Clinical Vignettas with Questions	Commitment to Change Without Follow-up	Commitment to Change With Follow-up	Patient Registries	Patient Registries (CareMeasures)	Patient Registries (CareMeasures)	Success Case Method; Interviews
Live Activities									
American Academy of Nurse Practitioners annual meeting (1)	453	X		X	X				
American Academy of Physician Assistants annual meeting (1)	447	X	X	X	X				
American College of Cardiology 2008 annual meeting (1)	225	X	X	X	X				
American Osteopathic Association state chapter meetings (21)	1,929	X		X	X				
Block Grants administrated through state medical societies (32 grantee organizations)	2,526	X	X	X	X				
Pharmacy counseling regional meetings (includes Webinar) (51)	2,400	X	X	X					
Primary Care Network (10)	4,486	X	X	X					
Primary Care Update (1)	375	X	X	X					
PriMed Conference and Exhibition (4)	1,000	X	X						
State Academy of Family Physicians (15) and American College of Osteopathic Family Physicians (2)	2,200	X	X						
Virginia regional hospital network (includes telemedicine) (11)	251	X	X	X	X				
Enduring Activities									
CardioSource (3)	72	X	X	X	X				
Discovery Health TV show and DVD (2)	979		X						
Epocrates (3)	8,039	X	X						
Learning from Self-Assessment (2)	512	X	X						
MedPage Today (1)	502	X	X	X					
Medscape (1)	4,472	X	X	X	X				
PriMed slide lecture series (1)	1,586	X	X	X					
ReachMD (1)	723	X	X						
Web-based cases on motivational interviewing (1)	16	X	X						
Performance Improvement									
Collaborative model-California Academy of Family Physicians (1)	60	X	X			X	X	X	X
Performance Improvement Workshop-Interstate Postgraduate Medical Association (1)	95	X			X				
Practice-facilitator model-Iowa Foundation for Medical Care (1)	28	X				X	X	X	X
Self-directed learning-University of Wisconsin (1)	99	X	X			X			X
Tools/Other									
Toolkit eResource via Skyscape	2,204	X							
Educational exhibits at national conferences (3)	785								
QuitAdvisor	154								
Web portal, including toolkit	6,389	X	X						X

Note: X = Outcomes were assessed

FIGURE 1. Tobacco Cessation Competencies Measured Using Clinical Vignettes: Relative Percent Difference Between Mean Competency Scores of CS2day Participants Post Activity and Comparison Group

The majority of participants attending pharmacy regional meetings who responded to the CTC questions post activity ($n = 779$; 45%) were likely or very likely to implement at least 1 practice change: the percentage varied from 68% to 75% across 5 categories of changes (ie, Ask, Advise, Recommend Medications, Counseling, and Refer). Participants of pharmacy webinars who responded to the CTC questions ($n = 268$; 41%) were less likely (39% to 45% across 5 categories of changes) to rate themselves as likely or very likely to implement changes.

The CTC approach required adjustment for specific activities. For example, in evaluation of the Medscape enduring activity, it was not technologically feasible to track individual responses and link them to individualized follow-up. Therefore, only the follow-up portion of the approach was

conducted, resulting in responses from 86 clinicians who reported a total of 222 changes implemented in their practice.

Performance outcomes of 3 PI activities, measured as compliance with 8 measures, are shown in TABLE 2. We combined data from 2 activities that used the same approach, where performance during the first quarter of the activity was compared with the end-of-project cumulative results reflecting performance throughout the project year. The third PI activity, reported separately, used a different approach, where baseline (ie, before-the-intervention) performance was compared with the end-of-project (noncumulative) performance. Mean results across measures within 1 activity and across different activities varied. We also observed wide ranges between minimal and maximum improvement on each measure in all activities. Detailed data on the individual PI activities

TABLE 2. Aggregate Performance Outcomes

| Measure	Starting Compliance, %	Ending Compliance, %	Change in percentage points
Collaboratory Model PI Activity (California Academy of Family Physicians)			
Combined with Practice-Facilitator Model PI Activity (Iowa Foundation for Medical Care):			
data from 36 practices/clinicians			
TOB-01: Ask about tobacco use	75.5	66.7	−8.8
TOB-02: Advise tobacco users to quit	69.6	73.9	4.4
TOB-03: Assess readiness to quit tobacco use	77.9	75.4	−2.6
TOB-04: Assist tobacco users who are willing to quit with a behavioral quit plan	79.2	80.8	1.7
TOB-05: Assist tobacco smokers who are ready to quit by recommending medication use	64.6	54.4	−10.2
TOB-06: Provide tobacco users who are NOT ready to quit with motivational treatment	57.9	70.9	13.0
TOB-07: Arrange follow up for tobacco users attempting to quit	34.5	29.8	−4.7
TOB-08: Assist former tobacco users with relapse prevention	0.0	50.0	50.0
Self-Directed Learning PI Activity (University of Wisconsin): data from 99 clinicians			
TOB-01: Ask about tobacco use	58.1	93.7	35.6
TOB-02: Advise tobacco users to quit	68.2	93.5	25.3
TOB-03: Assess readiness to quit tobacco use	54.2	90.4	36.2
TOB-04: Assist tobacco users who are willing to quit with a behavioral quit plan	82.7	99.8	17.0
TOB-05: Assist tobacco smokers who are ready to quit by recommending medication use	55.2	85.8	30.6
TOB-06: Provide tobacco users who are NOT ready to quit with motivational treatment	87.7	96.7	9.0
TOB-07: Arrange follow up for tobacco users attempting to quit	53.4	82.6	29.2
TOB-08: Assist former tobacco users with relapse prevention	62.5	91.3	28.8
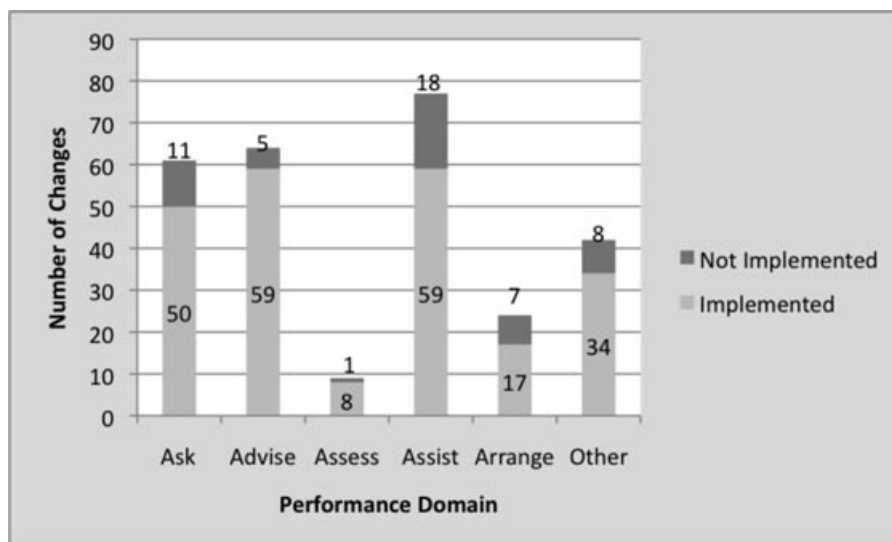
FIGURE 2. Aggregate Results From Live Activities: Changes Confirmed on Follow-up

are available in the article by Mullikin and colleagues[22] in this supplement.

Level 5—Patient Health Status, and Level 6—Population Health Status

In 2 PI activities that were evaluated at these levels, a change in smoking status had been recorded for 231 of 494 patients who were seen 2 or more times during the project, representing a quit rate of 46.8%.

SCM Study Results

From the SCM study of 9 outpatient practices, we concluded that in these instances, the PI activities were a primary and proximal cause of improvement in clinical practice. The mechanism through which the activity contributed to behavior change varied from helping learners develop new skills to providing practical tools, and depended on the clinical context, such as ability to customize electronic health records systems and prior experience using clinical data. See the article by Olson and colleagues[20] in this supplement for more detailed information.

Formative Evaluation and Continuous Improvement

Program planners utilized formative evaluation data to improve ongoing activities. For example, a CS2day session at the American College of Cardiology 2008 meeting drew a lower number of participants than expected, and in particular, a low number of US participants—47 (21%) of 225. In 2009, to better reach the target audience, the partners replaced this activity with an educational exhibit ($n = 192$ visitors) and an enduring CardioSource activity ($n = 72$ participants).

Another example was that the CS2day Web site was improved based on user satisfaction data. Also, 37 tools were updated one or more times based on new evidence or feedback from users/experts, and 7 videos were removed from the toolkit because of data pointing at controversial evidence and perceived commercial bias. Collaboration assessment identified several strengths but also areas for improvement that the partners acted on including program integration to reduce working in "silos" and improving partner access to evaluation data.

Discussion

We applied the outcomes-based CME evaluation model to guide evaluation[14] of the complex educational program. Below is our reflection on the results, followed by our insights into utilization of this model and the outcome assessment strategies we used in the context of the collaborative initiative.

The CS2day program had an overall positive impact on participating clinicians and their patients. The accomplishment of desired outcomes was documented across various program components at multiple levels of evaluation. A quit rate of 46.8% was achieved among smokers seen two or more times by clinicians participating in the PI activities. Although direct comparison with published results may be somewhat misleading due to differences in methods, we found that outcomes of previously reported multicomponent interventions varied between 7% and 15% of smoking abstinence in an intervention group.[23]

Based on prior publications,[23] we expected that the greatest need for improvement would be in providing assistance with quitting. Indeed, participants of the PI activities needed improvement in this area. However, notable performance gaps and the biggest improvements were observed among participants of one PI activity with respect to *ask about tobacco use* and *assess readiness to quit*, and among participants of two other PI activities with respect to *arrange for follow-up* and *assist for relapse prevention*. No single common pattern in terms of changes relative to the 5 A's competencies/performance domains became evident from assessments using clinical vignettes and the CTC approach. This suggests that practice gaps in tobacco cessation are likely to vary and future interventions to improve adherence to the evidence-based guideline[2] should target all 5 A's.

Observed differences in performance outcomes among PI activities may be partly explained by differences in educational designs and approaches to tracking performance.[22] These differences and wide ranges between minimal and maximum improvement on each measure also provide evidence of great variability in the impact of education on practice. The SCM study illuminated why and how educational intervention worked for those practices that achieved great improvement.[20] However, why did some participants/practices benefit so much and others so little from the project? Why did some participants decrease their performance on some measures? Conducting case studies of "successes" in comparison with "failures" may be a promising method to answer such questions in future evaluations.

CS2day outcomes were consistently tracked in every educational activity at multiple levels of evaluation but only 1 PI activity was evaluated at all 6 levels. Although there is strong rationale to conduct all-level evaluation in that analysis of several kinds and sources of data allows cross-validation and provides evidence related to several links in the probable chain of effects triggered by participation in education,[24] we did not evaluate all activities at all levels. We did not do it because it was not feasible, even for a sufficiently funded big initiative, or needed for each activity (eg, not needed for educational exhibits that were designed to primarily impact lower level outcomes). Also, on several occasions, we covered fewer evaluation levels than initially planned due to the activity design, technical problems, or lack of participants, which was the case in 1 PI activity.[22] Another reason to consider tracking selected rather than all-level outcomes is availability of clear evidence linking lower-level outcomes to performance and patient outcomes. We did not see this reason

applicable to smoking cessation,[25] but it may be a legitimate reason to simplify evaluation in some clinical contexts.[9,26]

Not surprisingly, we observed that higher-level evaluation is resource intensive and time consuming. Luckily, approaches like the CTC tool[27] and tracking/providing feedback on performance[28] are interventions in themselves capable of facilitating change, which adds value to their utilization and arguably increases return on investment. At the same time, the dual purpose of these approaches created a challenge in CS2day evaluation. From the educational perspective, tailoring of an intervention to the clinical context and participant needs was the appropriate thing to do. But from the evaluation perspective, such tailoring led to differences in sampling and data collection and other variations, making it harder to aggregate results.

Another observation relates to evaluation utilization. The CS2day partners used the results of formative evaluation in a timely manner, leading to many improvements in collaboration and ongoing education. Summative results were reflected upon by many stakeholders but not fully translated into implications for practice, and they were underutilized in planning the next phase of CS2day. Why did this happen? The Joint Committee on Standards for Educational Evaluation[29] recommended 7 best practices to increase evaluation utilization, and we did not follow 1 critical recommendation concerning *having a plan* to help the stakeholders assess, interpret, and apply findings from an evaluation report. Also, the richness of results—the 402-page-long final report and multiple supplementary reports—and insufficient time to reflect on the results due to the project's momentum and timing of funding opportunities for the next phase contributed to the utilization challenge.

Reflecting on the CS2day evaluation experience, we identified several major strategies that helped the partners overcome challenges associated with the evaluation complexity and conflicting perspectives of multiple stakeholders. First, utilization of 1 outcomes framework,[14] which the partners agreed on when planning the initiative, provided the structure for specifying desired outcomes and interpretation of results, and facilitated communication among program planners, faculty, project managers, and evaluators. Second, establishing common outcome measures and evaluation tools/techniques enabled consistent reporting, aggregate data analysis, and comparisons among different activities (although the latter was outside of the scope of this article). Third, to facilitate data organization and analysis, the partners utilized centralized registry databases for higher-level outcomes and a Web-based repository for the rest of the outcomes.

We also developed an understanding that collaboration synergy was critical in planning and implementing evaluation. Three conditions seemed to enable this synergy: the partners had complementary strengths (such as content expertise, skills in performance measurement, and proprietary tools), were willing to contribute their strengths, and were given opportunities to do so.

Lessons for Practice

- A multilevel outcomes framework is useful to facilitate agreement on evaluation purposes/methods among stakeholders with multiple perspectives and to guide the evaluation process.

- Establishing common outcome measures and evaluation instruments, and using centralized data repositories in a collaborative, multicomponent program enables consistent and appropriate assessment at each outcome level and aggregate data analysis.

- To understand how the program worked, it is important to complement outcomes assessment with evaluation methods examining mechanisms that link education to outcomes.

- A systematic effort to evaluate an educational program should include a systematic effort to utilize evaluation results starting with a plan for helping stakeholders to interpret and use an evaluation report.

The major limitations of CS2day evaluation included absence of cost-benefit analysis, lack of methods to assess sustainability of changes, and limiting statistical analysis to descriptive statistics. Other limitations were related to particular methods including having a comparison group in clinical vignettes assessment that consisted of physicians only and did not reflect the diverse specialties of CS2day participants, and using objective performance measures that were self-reported by participants, causing potential bias in patient sampling.

To conclude, outcomes assessment of a program like CS2day is a complex and multifaceted effort that requires consideration of multiple dimensions to develop a comprehensive understanding of the overall impact. We agree with Cronbach[30] that evaluation is an art, which means no single best evaluation plan can suit similar programs. However, some evaluation principles and good practices are important and replicable. Providers of continuing education who plan collaborative initiatives may benefit from the use of a multilevel, outcomes assessment framework, common measures and evaluation techniques, centralized repositories of data, and inclusion of evaluation methods designed to examine why and how the outcomes were achieved. Collaborative continuing educational programs have a great potential to improve health care and patient outcomes but more evaluation research is needed to document their impact and understand what program elements are most effective in achieving desired results.

References

1. Olson CA, Balmer JT, Mejicano GC. Factors contributing to successful interorganizational collaboration. *J Contin Educ Health Prof.* 2011;31(Suppl 1):S3–S12.

2. Fiore MC, Jaen CR, Baker TB, et al. *Treating Tobacco Use and Dependence: 2008 Update.* Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services, Public Health Service; 2008.

3. Kaufman DM. Applying educational theory in practice. *BMJ.* 2003;25(326):213–216.

4. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA.* 1999;282(15):1458–1465.

5. Walsh P. Evaluating educational activities. In: Adelson R, Watkins FS, Caplan RM, eds. *Continuing Education for the Health Professional: Educational and Administrative Methods.* Rockville, MD: Aspen; 1985:71–100.

6. Davis D, Barnes BE, Fox R, eds. *The Continuing Professional Development of Physicians: From Research to Practice.* Chicago, IL: American Medical Association; 2003.

7. Moore DE, Green JS, Gallis HA. Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. *J Contin Educ Health Prof.* 2009;29(1):1–15.

8. Tian J, Atkinson NA, Portnoy B, Gold RS. A systematic review of evaluation in formal continuing medical education. *J Contin Educ Health Prof.* 2007;27:16–27.

9. Belfield C, Thomas H, Bullock A, Eynon R, Wall D. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach.* 2001;23(2):164–170.

10. Marinopoulos SS, Dorman T, Ratanawongsa N, et al. *Effectiveness of Continuing Medical Education.* Evidence Report/Technology Assessment No. 149 (prepared by the Johns Hopkins Evidence-Based Practice Center, under Contract No. 290-02-0018). AHRQ Publication No. 07-E006. Rockville, MD: Agency for Healthcare Research and Quality; 2007.

11. Hager M, Russell S, Fletcher SW, eds. Continuing education in the health professions: Improving healthcare through lifelong learning. Proceedings of a conference sponsored by the Josiab Macy Jr. Foundation, Bermuda, November 2007. New York, NY: Josiab Macy Jr Foundation; 2008.

12. Cervero RM. Continuing professional education in transition, 1981-2000. *Int J Life Educ.* 2001;20(1–2):16–30.

13. Lasker RD, Weiss ES, Miller R. Partnership synergy: a practical framework for studying and strengthening the collaborative advantage. *Milbank Q.* 2001;79(2):179–205.

14. 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: American Medical Association; 2003:249–274.

15. Kirkpatrick DL. *Evaluating Training Programs.* San Francisco, CA: Berrett-Koehler; 1994.

16. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Ann Intern Med.* 2004;141:771–780.

17. Shershneva MB, Wang M, Lindeman GC, Savoy JN, Olson, CA. Commitment to practice change: an evaluator's perspective. *Eval Health Prof.* 2010;33(3):256–275.

18. Curry L, Purkis I. Validity of self-reports of behavior changes by participants after a CME course. *J Med Educ.* 1986;61: 579–586.

19. Wakefield J, Herbert C, Maclure M, et al. Commitment to change statements can predict actual change in practice. *J Contin Educ Health Prof.* 2003;23(2):81–91.

20. Olson CA, Shershneva MB, Brownstein M. Peering inside the clock: using Success Case Method to determine how and why practice-based educational interventions succeed. *J Contin Educ Health Prof.* 2011;31(Suppl 1):S50–S59.

21. Mattessich PW, Murray-Close M, Monsey BR. *Collaboration: What Makes It Work.* 2nd ed. St Paul, MN: Amherst H. Wilder Foundation; 2001.

22. Mullikin EA, Ales MW, Cho J, Nelson TM, Rodrigues SB, Speight M. Sharing collaborative designs of tobacco cessation performance improvement CME projects. *J Contin Educ Health Prof.* 2011;31(Suppl 1):S37–S49.

23. Papadakis S, McDonald P, Mullen KA, Reid R, Skulsky K, Pipe A. Strategies to increase the delivery of smoking cessation treatments in primary care setting: a systematic review and meta-analysis. *Prev Med.* 2010;51:199–213.

24. Knox AB. *Evaluation for Continuing Education: A Comprehensive Guide to Success.* San Francisco, CA: Jossey-Bass; 2002.

25. Unrod M, Smith M, Spring B, DePue J, Redd W, Winkel G. Randomized controlled trial of a computer-based, tailored intervention to increase smoking cessation counseling by primary care physicians. *J Gen Intern Med.* 2007;22(4):478–484.

26. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ.* 2000;321:694–696.

27. Pereles L, Lockyer J, Hogan D, Gondocz T, Parboosingh J. Effectiveness of commitment contracts in facilitating change in continuing medical education intervention. *J Contin Educ Health Prof.* 1997;17: 27–31.

28. Grol R, Wensing M, Eccles M. *Improving Patient Care: The Implementation of Change in Clinical Practice.* Toronto, Ontario: Elsevier; 2005.

29. Sanders JR (Chair), the Joint Committee on Standards for Educational Evaluation. *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs.* 2nd ed. Thousand Oaks, CA: Sage; 1994.

30. Cronbach LJ. *Designing Evaluations of Educational and Social Programs.* San Francisco, CA: Jossey-Bass; 1989.