As outcome measurement becomes an increasingly critical adjunct to learning activities for continuing medical education (CME), both supporters and providers need cost-effective ways to include robust outcome measurement not only of learning gains (Kirkpatrick Level 2) but also of corresponding changes in clinical practice (Level 3) and patient outcomes (Level 4). Healthcare providers (HCPs) are often surveyed and asked whether their clinical practices will change or have changed as the result of the educational activity in which they have participated. Because this type of self-reported survey information is not based on direct measurement or observation of actual behavior, it is not outcomes data per se but is rather the respondent's opinion about past or future outcomes. As such, it is not a "measurement" but is an informed guess.

In spite of these inherent limitations, the self-reported survey is the only method commonly used to assess the clinical impact of a CME program, because the current methods for assessing changes in clinical practice and patient outcomes by means of direct observation are simply too costly and time-consuming for all but the occasional well-funded research project. These include the standard patient method (to assess clinical practice), which uses trained actors and actresses posing as real patients; and the chart analysis method (to assess patient outcomes), which typically involves review of paper charts and hand-keying of relevant data.

The interactive clinical case study simulation offers an extremely cost-effective method for measuring the impact of a CME activity on the learner’s clinical practice. Although it cannot directly observe the HCPs behavior in the actual clinical setting, it can directly measure the choices the HCP makes in diagnosis, treatment, and follow-up. By simulating the most common patient profiles for a particular disorder, the interactive clinical case study can record the HCPs diagnostic and treatment decisions as he or she works through the case, show the likely outcome of each decision, and provide individualized feedback to identify and correct inappropriate decisions.

Case studies have long been used in medical education, and their effectiveness as an instructional strategy is well established. In fact, clinicians’ scores on case study exercises have been found to be a more accurate reflection of actual clinical practice than traditional chart analysis [1,2]. Interactive simulations of clinical case studies can be created cost-effectively by loading specific case information into a pre-programmed software template that has been designed to mimic, in generic form, the clinical diagnostic and treatment process. Such simulations can be constructed as text-only modules for maximum cost effectiveness, as multimedia modules with audio, video, and animation for maximum educational impact, or as an intermediate blend of text and multimedia. Whatever the level of production, the interactive clinical simulation can be readily programmed to present a typical patient profile and then record every decision the learner makes in diagnosis, treatment, and follow-up. Administered prior to an educational program or event, the simulation can record the clinician's baseline skills and knowledge. Administered again after the educational event, a parallel simulation can measure the learner’s improvement in making diagnostic and treatment decisions—in other words, the impact of the educational event on the learner’s clinical thinking.

In a more complex form of the interactive case study simulation, the learner is presented with feedback on the consequences of each clinical decision, given the opportunity to revise diagnostic and treatment decisions, and guided toward the optimal diagnosis and course of treatment. This type of simulation goes beyond the simple assessment of clinical competence and provides an individualized educational experience targeted at precisely those areas in which the learner needs remediation.

Modern computer technology has made possible the creation of simulated clinical experience and the precise measurement of patterns of clinical decision-making within a technology environment that can deliver quantitative outcomes data rapidly and at a lower cost than any other outcome measurement strategy. Because the interactive case study directly measures learner behavior, it generates data of intrinsically greater validity than the self-administered survey questionnaire; and because it can be distributed and administered entirely by computer, it can do so with a minimal expenditure of time and resources.
INTRODUCTION:
With the implementation of the new accreditation guidelines of the Accreditation Council on Continuing Medical Education (ACCME) [7], it has become clear that CME providers will be required to integrate outcomes measurement into the planning and execution of accredited medical educational activities. Heretofore, outcome measurement has been limited generally to measuring participation (ie, the presence or absence of warm bodies) and acceptance (ie, the learners' degree of satisfaction with the quality and value of the program). But funding organizations know that it is possible for learners to participate in an educational activity, have a generally good opinion of it, and yet learn little that is new and even less that results in the desired changes in clinical care.

The newly revised guidelines for maintaining CME accreditation announced by the ACCME will take effect in January 2008. They include a requirement that medical education providers show progress toward developing educational strategies that incorporate “higher levels” of outcome measurement. These include measurements of the amount of learning that took place as a result of the activity (Level 2), the changes in clinician behavior that are directly attributable to that particular learning experience (Level 3), and the changes in patient outcomes resulting from the desired changes in patient care (Level 4).

At the same time, supporting organizations are understandably reluctant to support the addition of expensive and time-consuming outcome measurement studies to verify that the event they supported accomplished more than attracting a number of warm bodies who are willing to say that they thought the program was valuable. These stakeholders want and need to know what the impact of their program will be on the quality of care, and how that impact will be measured.

Before discussing specific solutions, such as the interactive clinical simulation, it will be necessary to review a contemporary version of Kirkpatrick’s scheme to ensure that the terms used in this article are properly understood by all readers.

LEVELS OF OUTCOME MEASUREMENT

The preferred conceptual scheme for measuring learning outcomes is based on the pioneering work of Donald Kirkpatrick [3,4], who developed a formulation that over time has become the foundation upon which most outcome measurement approaches are based. Kirkpatrick’s original scheme identified 4 distinct levels of outcome measurement:

Level 1: “Reaction” (the extent of learners’ participation and satisfaction)
Level 2: “Learning” (the amount of knowledge absorbed by learners)
Level 3: “Behavior” (the degree to which the learner’s behavior has changed)
Level 4: “Results” (changes in the learner’s productivity and effectiveness)

Kirkpatrick’s great contribution to outcome measurement was not only his recognition that the outcomes of educational programs had to be measured on several distinct levels, but also that the validity of the data gathered at each level depends on the validity of the data gathered at the level below it [2,4]. For example, it is obvious that effective learning cannot occur unless learners react positively to an educational program and give it their active attention. By the same token, changes in clinical behavior can only be attributed to an educational event if the educational outcomes of that event have been measured and can be correlated with these changes.

For this reason, Kirkpatrick’s scheme is usually represented graphically in the form of a triangle or pyramid, with Level 1 as the base and Level 4 as the apex, each level resting on the level beneath it. Eventually, a fifth level was added to Kirkpatrick’s 4-level scheme by Jack Philips of the ROI Institute [5]. Phillips’ fifth level represented the educational program’s net economic benefit to the learner’s entire organization. In the case of medical education, however, Level 5 is usually represented as the health care benefits to an entire patient population.

Figure 1 is a typical representation of the 5 levels of outcome measurement, illustrating the principle that the validity of the outcomes data gathered on each level rests upon the valid measurement of the outcomes data on the level beneath it.

- Level 1 outcome measurement (participation and satisfaction) is already standard practice in CME and constitutes the bulk of the CME outcomes that are actually measured.
- Level 2 outcome measurement (learning gains) occurs in many CME activities,
although it is far from standard practice. Simple text-based pre/post and follow-up tests, however, are relatively easy to create and administer, and the process of identifying and measuring educational gains made between pre-test and post-test is straightforward.

- Level 3 outcome measurement (changes in clinical behavior) is arguably the most difficult type of outcome to measure, yet at the same time its measurement is critical to establishing a causal link between the educational intervention and subsequent improvements in patient outcomes.

- Level 4 (changes in patient outcomes) and Level 5 (changes in population outcomes) will continue to be time-consuming and expensive to measure in healthcare environments using traditional paper charts. However, those organizations that have instituted an electronic health record (EHR) system will be able to quickly and easily identify and quantify specific improvements in patient outcomes and their corresponding cost savings. But these improvements are linked definitively to specific educational events only if both the learning gains and the changes in learners' clinical behavior are measurable and quantifiable. Pre-/posttesting can effectively measure the learning gains resulting from the educational experience. The interactive case simulation can effectively measure the changes in clinical decision-making resulting from the educational experience.

The interactive clinical case study simulation yields scientifically valid data because it directly observes and records clinical decision-making in real time. Although it does not provide direct observation of the learner’s actual clinical practice, it provides a much firmer foundation for making inferences about actual clinical decisions—more than the undocumented and unsupported opinions gathered by typical survey questions.

MEASURING CHANGES IN CLINICAL PRACTICE

Due to both its high cost and daunting legal considerations about privacy, no affordable method of direct observation is available to gather data on the HCP’s actual clinical behavior toward actual patients. However desirable in an ideal sense, the outcomes researcher in the real world cannot employ some version of the proverbial “fly on the wall” to make direct observations of clinical practice in the examining room.

The most valid direct method of observing changes in clinical practice is the “standard patient” method, in which professional actors and actresses are trained to describe the symptoms of a specific disorder or disease state and pose as real patients with scheduled office visits, their identity known only by the research team, and unknown to the clinician. The standard patient role requires not only training the actor or actress to present a convincing portrayal of a particular patient type, but also training them to prepare a detailed report on the nature and quality of the clinical care received. While this method rests on direct observation and therefore has the greatest scientific validity, it is slow, extremely labor-intensive, and far too costly to be used as an outcome measurement strategy for more than the occasional high-end research study.

Self-Administered Surveys

The most common indirect measurement of clinical behavior change is the self-administered survey method. Although it is cost-effective and produces quantitative data, the survey method has significant weaknesses as an assessment tool.

Although it can be reasonably assumed that clinicians are answering survey questions as honestly and faithfully as possible, these data are essentially the learner’s opinion of changes in his/her clinical practice, and is based on memory and subjective impression rather than on facts systematically recorded. There is no way of determining whether self-reporting is accurate or inaccurate in any particular case. For these reasons, data gathered by post-event self-reporting are assessments not of changes in clinical practice, but rather of how learners believe their practices have changed or will change in the future. Thus, although self-administered surveys gather opinions, they do not directly measure anything.

The Print-Based Case Study

Print-based case studies have been used as teaching strategies in medical education for a very long time. Traditionally, they have consisted of descriptions of specific types of cases or patient profiles with relevant history and physical information, followed by multiple-choice questions about which diagnostic and treatment procedures the learner would recommend. This approach has proven educationally effective with clinicians because they are by nature problem-solvers, and also because problem-solving is one of the most effective ways of learning, especially for adults. Clinicians’ responses to the problems posed by case studies have also proven to be reliable and effective indicators of the quality of actual clinical care [1,2]. However, the print-based case study has serious limitations: it cannot automatically record every clinical choice made by the learner, and it cannot provide feedback to the learner on either the appropriateness or the consequences of the various clinical decisions made in the course of diagnosis and treatment.
The Interactive Case Study Simulation

It is not difficult to create an interactive software module designed to simulate the diagnostic and treatment phases of the actual clinical process. As the learner progresses through the case, a variety of diagnostic and treatment choices are presented, and by means of the interactive technique known as “adaptive branching,” the learner’s choices determine—within reasonable constraints—the clinical progress of the case. At the same time, the software module is tracking every clinical choice and responding to those choices by branching to content that reflects the consequences of each choice. These interactive features contribute significantly to enhancing learning effectiveness, even as assessment is taking place. In short, the software-driven case study simulation functions both as an assessment instrument and as a problem-based learning environment.

Table 1 illustrates the types of data that would be captured during the course of a typical interactive case simulation. At each step in the diagnostic and treatment process, the learner is required to make choices. Which elements of the presenting complaint and the patient’s history and physical are relevant to the case? Will the appropriate diagnostic tests be ordered and the results correctly interpreted? Will the learner pursue the most appropriate treatment strategy? Will the learner correctly interpret the patient’s response to the treatment in follow-up visits? Will the learner adjust the treatment strategy correctly when problems are identified in follow-up?

The Validity of Level 3 Outcomes Measurement: The Case Study Simulation versus the Self-Administered Questionnaire

Although it cannot measure the learner’s actual clinical behavior, the case study simulation can directly measure the learner’s clinical decision-making patterns. By tracking the learner’s behavior in the case study simulation before the educational activity, baseline data on decision-making patterns can be collected. These data can then be compared to the results of follow-up scenarios presented after the educational experience, using a parallel mix of patient profiles.

Most importantly, the Level 3 outcomes data generated by the case study simulation has, for several reasons, inherently greater validity than data gathered by conventional self-reporting survey instruments.

First, the simulation model is closer in format to the actual clinical experience. Rather than asking the physician to supply cognitive estimations of the effect that the educational experience has had (or will have) on clinical practice, the case study simulation measures the actual decision-making process of the clinician when confronted by a real-world clinical problem. Thus, data gathered by the case study simulation present a firmer foundation for inferring what the clinician is doing in actual practice, than self-reported subjective data based on untestable mental impressions.

Second, the case study simulation can be programmed to gather data that are rich in nuance and detail. It can measure the length of time each learner took to make each clinical decision and will know whether the learner hesitates and makes changes in a particular diagnostic or treatment decision or, by contrast, proceeds immediately to a choice and moves on. The simulation’s underlying technology can potentially record every mouse click, allowing complex outcomes measurement strategies to be designed for a wide variety of educational applications.

Third, case study simulations can be designed using a “guided discovery” strategy designed to prevent the learner from wandering off in unproductive directions and to keep the clinical process moving in the direction of an appropriate resolution. Thus, the learner is not allowed to proceed to a definitive diagnosis without first performing or ordering the essential diagnostic procedures or tests. Similarly, the learner cannot proceed to the selection of a treatment strategy without first set-

<table>
<thead>
<tr>
<th>CLINICAL DECISIONS</th>
<th>DATA CAPTURED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selective review of presenting complaint, history, and physical information</td>
<td>Those aspects of the H&amp;P that are most likely to be reviewed, and in what sequence</td>
</tr>
<tr>
<td>Selection of diagnostic tests for relevance</td>
<td>Tests most likely to be selected as relevant</td>
</tr>
<tr>
<td>Selection of diagnostic tests to order for this case</td>
<td>Probability that each of the relevant tests will be ordered</td>
</tr>
<tr>
<td>Selection and ranking of differential diagnoses</td>
<td>Probability of each of several possible diagnoses being selected</td>
</tr>
<tr>
<td>Selection of definitive diagnosis</td>
<td>Most likely diagnosis to be selected</td>
</tr>
<tr>
<td>Selection and ranking of treatment strategies</td>
<td>Probability of each of several possible treatment strategies being selected</td>
</tr>
<tr>
<td>Selection of a definitive treatment strategy</td>
<td>Most likely treatment strategy to be selected</td>
</tr>
<tr>
<td>Follow-up and revision of treatment strategy</td>
<td>Revision of treatment strategy as needed based on the results of follow-up visits</td>
</tr>
</tbody>
</table>
tling on the correct definitive diagnosis. Having found the correct diagnosis, the learner must then select the most appropriate treatment strategy and respond appropriately to problems or complications that are identified during follow-up.

At each of the decision points in the clinical simulation, the case study simulation gathers data about the physician’s preferences, tendencies, certainties, and uncertainties. Together with demographic data captured at registration, the resulting outcomes data are rich in information about the changes in clinical decision-making brought about by an educational program or event. Because these data are already in digital form, they can be analyzed for Level 3 outcomes as well as for statistical significance in a matter of hours.

Because the time and expense required to measure clinical behavior is often prohibitive, the case study simulation provides a sophisticated yet cost-effective assessment instrument, from which actual clinical behavior patterns can be inferred with a greater degree of confidence than by any other means currently available to the medical educator.

REFERENCES

7. The ACCME Essential Areas and Their Elements, Accreditation Council for Continuing Medical Education, Chicago, IL, August 2007.