Evaluation of the OncoTCap (Oncology Thinking Cap) Simulation in Teaching Medical Students About Clinical Trials: Some Successes and Some Surprises

ROGER S. DAY, SCD, BRENDA K. MANNING, PHD, DENISE GARROTT, PHD, ADAM M. BRUFSKY, MD, PHD, JOSEPH BAAR, MD, PHD, MERRILL J. EGORIN, MD, PHD, WILLIAM E. SHIREY, MS, CHARLES FRIEDMAN, PHD

Abstract—Objective. A training workshop for cancer clinical trials was developed utilizing a computer simulation encompassing cancer biology, metabolism, adverse effects, and clinical trial design. Method. Fourth-year medical students in a neoplasia elective course participated. The workshop was structured to maximize group discussions and interactions. Pretests and posttests were administered in a crossover design to evaluate learning about cancer clinical trials. Results. The comparison showed that the workshop did impart knowledge about cancer clinical trials. Conclusions. Student evaluations of the workshop showed slight improvements from the first year to the second year. J Cancer Educ. 2004; 19:149-155.

Medical students with a special interest in oncology are usually instructed through traditional lectures and supplemented through direct experiences with instructors in clinical and laboratory settings. Direct experiences appeal to students anxious to move forward in their medical careers. Arguably, in comparison to lectures, they can transmit a deeper understanding, perhaps with more lasting impact. However, experience-based education is limited to setting. For example, a medical student will not be able to observe the life cycle of a real clinical trial in the short time available. Also, some concepts used in drug development related to drug resistance are related to cancer cell population dynamics, for which no direct experience is readily available. For these reasons, computer simulation can be a useful adjunct to cancer education of medical students. It provides settings for experiences, albeit artificial, that cannot be provided directly. Simulation in medical education has focused primarily on acquiring specific skills rather than concepts. In contrast, this report describes a computer-simulation workshop for 4th-year medical students focused on concepts of clinical trials instead of skill acquisition.

The OncoTCap (Oncology Thinking Cap) cancer simulation software has multiple uses. The workshop described in this article teaches 4th-year medical students about complexities and interdependencies in clinical trials design, the role of clinical trials in the drug development process, and the eventual availability of new drugs for acute care and long-term management. While few of the students are expected to become clinical researchers, the workshop can prepare nonresearcher physicians to be more informed consumers of research reports. We also hope to encourage medical students to consider clinical research in their career plans.

We report results from a battery of evaluation strategies designed to assess the nature of students’ interactions with OncoTCap (process), their affective responses to the computer program and presentation format (reaction), and scores on pre- and posttests (learning).

METHODS

Learner Characteristics

All participants were 4th-year medical students at the University of Pittsburgh School of Medicine. In 2001, 25 students were enrolled, of whom 18 participated in the workshop; in 2002, 17 students were enrolled, all of whom participated. Initially, few students in either year declared or dem-
onstrated knowledge and experience in statistics or medical research; however, most students reported “advanced” computer skills. Respectively in 2001 and 2002, 12 of 18 and 12 of 17 students reported that they envisioned either “full-time research” or “research and practice” as a career. “Match Day” occurred near the end of the course, making the faculty’s task of keeping the students’ focus engaged especially challenging.

The OncoTCap Workshop

OncoTCap Version 3 provides an extensive and highly flexible cancer modeling laboratory enabling a range of educational interventions. Its development is supported by the Educational Resource for Tumor Heterogeneity (ERTH) project. The software is written in the Java computer language and runs on Windows, Unix, and Macintosh computers. Educational and research scenarios are authored in the Protégé knowledge management system. Upon deployment, each scenario appears in a generic interface providing introductory screens, a decision-scaffolding screen, and screens for displaying and interacting with simulation output.

OncoTCap was incorporated as a curriculum element into Neoplasia and Neoplastic Diseases (MED 5175), a month-long intensive course designed for 4th-year students, in March 2001 and March 2002. In addition to interacting with OncoTCap, students attend lectures, a clinical rotation, and journal club to meet selective requirements. An initial lecture on Biostatistics providing an orientation to Phase I, II, and III trials is a standard element in the MED 5175 curriculum.

The lecture is augmented with simulations based on decks of cards allowing students to role-play as patients and as principal investigators. The scenario utilized in the workshop described here centers on design and execution of a Phase II clinical trial. The drug under development in the scenario is a fictional anticancer agent. Its properties were chosen to reflect agents in actual clinical use. These properties include mechanisms of action, mechanisms of resistance, pharmacokinetics, and toxicity profiles. Accurate predictive modeling is well beyond the scope of this exercise. Therefore, to prevent overinterpretation, a fictional name (in this case “Brufomycin”) is assigned to the investigational drug.

The learning objectives of the workshop are as follows:

- To experience the decisions and dilemmas of clinical trial design by writing a protocol,
- To utilize clinical and preclinical pharmacokinetic information for dose planning,
- To appreciate the effects of chance on clinical trials by observing results of clinical trial replications,
- To observe “clinical response” as a phenomenon of tumor dynamics and appreciate the uses and limitations of clinical response as a measured surrogate endpoint, and
- To experience the effects of patient variability on patient management and clinical trial outcomes.

The Basic Activity of the Workshop: Clinical Trial Design and Execution

Students begin the exercise by reviewing a sequence of screens representing the introduction and objectives sections of a protocol. The sections are marked up with hyperlinks to explanatory documents and with highlighting of sections for future reference. The next screen (Figure 1) instructs the students to complete the protocol by selecting statements for inclusion from a list of 10 statements. Most statements also require students to make “fill-in-the-blanks” selections, for example, for treatment regimen details, statistical design, or dosage reduction rules. Some of these statements, such as treatment regimen and study design, are required. When all required statements are selected and complete, the clinical trial can be run.

Presentation Strategies

In 2001, six groups of 3 to 5 students worked on OncoTCap on a single laptop placed in an ERTH project conference room. These groups were facilitated by an educational specialist (DG). The scheduling was difficult and the time commitment was onerous. Therefore, in most sessions no oncologist or cancer researcher was present, a clear deficiency.

For these reasons, workshop presentation details were redesigned for the following year. In 2002, the OncoTCap program was installed in the student computer lab. Students came to the lab in two groups of 8 to 10 students. The program had been installed on individual workstations, each one shared by a pair of students. The ERTH principal investigator (RD) served as facilitator in this setting. To begin, a printout of a computer screen was distributed to each student pair (Figure 1), showing the configuration of a particular completed clinical protocol. The students, using the printout as a guide, modified their own running computer program screen. This activity led them to follow the steps of developing a series of protocol sections for a single agent Phase II anticancer drug trial, including eligibility criteria, treatment plan, patient management rules, data collection, and statistical design. By setting up this protocol, students became familiar with the computer program’s interface. Next, each pair of students ran a simulated clinical trial. The pairs reported their outcomes in terms of sample size, number of responses, and the study conclusion. By reporting, tabulating, and comparing these outcomes, the class obtained a sense for the effect of pure chance on the variability of clinical trial results. A sample of the output screen is seen in Figure 2.

The next step was to distribute a different screen shot to each pair of students, each screen shot differing from the first protocol in a single element. Student pairs set up the new clinical trial design and ran the corresponding simulated trial. The changed elements included removing the eligibility criterion for adequate liver function; adding an eligibility criterion restricting entry by race; increasing the number of
courses and doses per course; including a dose reduction rule to manage toxicity; increasing the dose past the Phase I MTD; including pharmacokinetic measurements; and taking a patient off-study for toxicity, new metastases, or progression. Each pair of students reported verbally to the rest of the workshop participants on the change made and the differences in outcome. The student reports provided opportunities for lively discussions on a host of issues, both scientific and social. For example, one student pair received a screen shot in which a race-restricted eligibility criteria was added, ostensibly to reduce variability in metabolic enzymes and improve the scientific validity of the study. When the simulation was run, the simulated protocol was disapproved by a simulated Institutional Review Board, providing an occasion to discuss the evolution of the scientific community’s views of race, sex, and age restrictions.

Subsequent exercises focused attention on the statistical aspects of study design, the planning and interpretation of pharmacokinetics studies, and the interpretation of clinical adverse events in relation to protocol protective elements such as dosage modifications. A “peek under the hood” in the output screen showed the tumor growth curves of each subpopulation of tumor cells, engendering a discussion of drug resistance mechanisms and the nature of clinical response as a surrogate endpoint. The final stage was open-ended, allowing students to make their own clinical trial design decisions, run their own trials, make their own observations, and share their experiences with the other workshop participants. The entire workshop lasted two hours.

Evaluations

The evaluation plan, outlined in Table 1, contained multiple elements designed to gather both qualitative and quantitative data about the impact of OncoTCap on the student as learner. Three evaluations were performed: process (how students learned), reaction (students’ affective response) to the OncoTCap experience, and learning (students’ cognitive gain).

Design of the Process Evaluation

The purpose of the process evaluation was to improve the computer interface, the workshop content, and the educational format. Evaluation experts recommend qualitative methods as well-suited for assessing process in medical education. We utilized videotapes of student interactions with the simulation and written comments from an exit questionnaire. In 2001 only, informal feedback was obtained immediately after each group session and a structured group interview was conducted with a sample of students drawn from the entire class at the end of the workshop.
FIGURE 2. Screen shot of a simulated clinical trial outcome. From top to bottom: study design, a study summary table, individual patient data (with tabs for clinical events, concentrations of drug and metabolites, tumor cell burden by type of cancer), and study conclusion. Here the tumor burden graph tab is shown.

TABLE 1. Sources of Qualitative and Quantitative Evaluation Data

<table>
<thead>
<tr>
<th>Evaluation Element</th>
<th>Used in 2001</th>
<th>Used in 2002</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Videotapes of group meetings</td>
<td>Yes</td>
<td>Yes</td>
<td>Provided qualitative data about how groups and individuals interacted with OncoTCap and the facilitator. Valuable for analyzing how students learned from the simulation, group process, and interaction with the expert facilitator. Tapes were analyzed by noting questions and statements that reflected students’ thinking and decision making.</td>
</tr>
<tr>
<td>Written evaluation questionnaire</td>
<td>Yes</td>
<td>Yes</td>
<td>Provided qualitative and quantitative data about students’ self-assessed knowledge and attitudes and plans for medical career. Valuable for comparing 2002 to 2001 self-assessed knowledge and attitudes.</td>
</tr>
<tr>
<td>Immediate postsession informal feedback</td>
<td>Yes</td>
<td>No</td>
<td>Provided qualitative data about students’ immediate responses to OncoTCap program, group process, and interaction with facilitator.</td>
</tr>
<tr>
<td>Structured composite group exit interview</td>
<td>Yes</td>
<td>No</td>
<td>Performed in 2001 to get feedback from a representative sample of all the groups regarding the value of OncoTCap as a learning tool. Valuable for gathering data across all groups. In 2002 we implemented changes to address major issues raised in this interview in 2001.</td>
</tr>
<tr>
<td>Pre- and postintervention multiple-choice tests</td>
<td>No</td>
<td>Yes</td>
<td>2002 only. Measured learning outcomes.</td>
</tr>
</tbody>
</table>
Design of the Reaction Evaluation

Students filled out a brief questionnaire after their last interaction with OncoTCap, concentrating on student satisfaction (reaction to the simulation as a learning experience). The questionnaire contained 10 questions eliciting affective responses about the workshop experience on a five-point Likert scale, with an additional 10 open-ended questions. Results for the 2 years were compared using Fisher exact tests, after collapsing the top two categories ("Agree" and "Strongly Agree") and the bottom three categories.

Design of the Learning Evaluation

In 2002, a quantitative evaluation of learning outcomes (understanding of clinical trials and drug development) during the OncoTCap workshop was designed based on administration of brief preintervention and postintervention tests. To avoid training and cuing effects solely from exposure to the questions, different questions were given before and after the intervention. Students were trained in two separate groups, labeled A and B. There were two groups of questions, labeled SP and YB (named for the colors of the question sheets). The design scheme and sample sizes are in Table 2. With this evaluation design, no students saw the same question twice, and the questions were roughly balanced with respect to their appearance before ("Pre") versus after ("Post") the workshop.

The students' answers were graded subjectively by a research oncologist (JB) on a 1 to 10 Likert scale. Because the analysis required that scores would be ranked later within each question, the grader was asked to focus primarily on assuring that the answers were correctly ranked as to quality and to be less concerned with the absolute meaning of a score.

Linear models for the raw scores and the ranked scores were fitted using least-squares. The pre-versus-post effect was tested controlling for student and for question. In addition, a randomization test was performed, with the observed proportion of rank pairs concordant with improvement ("post" better than "pre") as the test statistic. The period (pre/post) was permuted randomly within each set of answers, and the proportion of concordance recalculated on the reconstructed data set. This was repeated 5000 times to provide a null reference distribution for the proportion of concordance. Of the 20 students participating in the workshop, 14 were present for both the pretest and posttest, 1 arrived after the pretest, and 5 had to leave before the posttest was given. Results include all students unless otherwise noted.

RESULTS

Process

Videotapes of group sessions in 2001 showed that students quickly became actively engaged with the OncoTCap simulation and participated in "thinking out loud" and discussing what data to input. Through these behaviors, they exhibited problem solving and critical thinking regarding decisions about the complex set of variables that constitute a clinical trial protocol. Informal student feedback at the end of each group session included positive comments related to the workshop's learning objectives, including "It gives you a realistic picture or 'sped-up' view of clinical trials," "A good sense of the clinicians' balancing act between efficacy and toxicity," "Appreciation of patient-to-patient variability," "Difficulty of bringing good drugs to patients," and "Complexity of trial design and outcomes related to dose planning and toxicity." However, results of the structured group exit interview suggested that the students perceived that they would have been able to learn more from OncoTCap if they had a better mastery of statistics concepts and had access to additional options in the program (including a richer set of patient inclusion/exclusion criteria, more dosing options, and the option to treat toxicity rather than change the dose or stop treatment). In addition, several students mentioned that they would have liked to be able to discuss OncoTCap output with a research oncologist or biostatistician "in order to get the most out of it."

In response to these comments, several changes were made for 2002. Changes in presentation strategies have been discussed previously. In addition, a set of card games was developed, one game each for Phase I, II, and III clinical trials. The Phase I game was used primarily to demonstrate design issues and ethical issues in Phase I trials. The Phase II game was used primarily to demonstrate the meaning of Type I and II errors as well as the operation of stopping rules. The Phase III game was used primarily to demonstrate the meaning of P values. The concrete nature of the exercise was expected to provide experiential associations for the targeted concepts that would make them more meaningful and memorable. No formal evaluation was done, but each year several students provided unsolicited reactions to these games that were uniformly positive. Changes in the program itself included developing and hyperlinking supportive information on

<table>
<thead>
<tr>
<th>Questions SP1-SP4</th>
<th>Questions YB1-YB4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period</strong></td>
<td><strong>No. of Students</strong></td>
</tr>
<tr>
<td>Group A</td>
<td>Post</td>
</tr>
<tr>
<td>Group B</td>
<td>Pre</td>
</tr>
</tbody>
</table>

Table 2. Design of a Quantitative Cognitive Evaluation
pharmacokinetics, how Phase I starting doses are chosen, and assessment of clinical response and toxicity.

**Reaction**

Results of the questionnaire are displayed in Table 3. The largest change from year to year was a large increase in the proportion who found the printouts to be useful \((P = 0.007)\). This change is misleading, however, as the 2002 students probably had in mind the screen shots of completed designs (e.g., Figure 1), not available in 2001. No other differences were statistically significant. The next largest magnitude change was an increase in the proportion finding the simulation to be a valuable part for the course (50% to 65%). The changes in the method of involving the facilitator seemed to be an improvement.

**Learning**

The brief questionnaire contained open-ended questions that asked about students’ learning in relation to five of the seven learning objectives. Each question used the stem “what was the most important thing you learned about …?” Students’ responses demonstrated that they were able to express significant learning of concepts related to clinical trials (Table 4).

**Pre/Post Test Results**

Comparisons of pre- versus postintervention results can be seen in Figure 3. Note that for all questions except YB-2 and YB-3 there was a moderate to large increase in the score. Linear modeling controlling for student and question revealed that the pre-versus-post effect was highly significant for both the raw score \((P < 0.0001, F = 29.35)\) and the ranked score \((P = 0.0035, F = 9.02)\). The observed proportion of rank pairs concordant with improvement (“post” better than “pre”) was 67%. Out of 5000 such permutations, 54 were higher \((P = 0.011)\). The responses to SP-2 and SP-3 were identical; therefore, the analysis was repeated excluding SP-3, yielding \(P = 0.016\).

**Table 3. Student Affective Responses**

<table>
<thead>
<tr>
<th>Evaluation Question</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>I learned a lot about designing Phase II trials using the computer simulation</td>
<td>83</td>
<td>88</td>
</tr>
<tr>
<td>The computer simulation was easy to use</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>The computer simulation was easy to understand</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>The facilitator(s) were helpful in learning how to use the software</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>The facilitator(s) helped the group to think about the trial design</td>
<td>82</td>
<td>94</td>
</tr>
<tr>
<td>The clinical trial simulation was a valuable part of the Neoplastic Diseases course</td>
<td>50</td>
<td>65</td>
</tr>
<tr>
<td>I needed more direction or information in the initial trial design (before using the computer simulation)</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>The computer simulation printouts were helpful</td>
<td>22</td>
<td>71</td>
</tr>
<tr>
<td>The individual patient summary was easy to understand</td>
<td>56</td>
<td>76</td>
</tr>
<tr>
<td>I would like to have spent more time using the software</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

**Table 4. Prompted Recall Questions and Representative Responses**

<table>
<thead>
<tr>
<th>Question About</th>
<th>Representative Student Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose planning (Learning Objective 2)</td>
<td>Has an important impact on toxicities as well as clinical response. Thus, must tailor dose to achieve optimal levels of both. It can be hard to get a doable regimen [sic] with acceptable toxicity. More is not necessarily better.</td>
</tr>
<tr>
<td>Statistics (Learning Objective 3)</td>
<td>They can allow us to show evidence of effectiveness or toxicity. Statistics directly determine your trial size and how you can calculate whether a drug is or is not promising.</td>
</tr>
<tr>
<td>Patient toxicity (Learning Objective 5)</td>
<td>There are many important variables in toxicity—dose, duration, individual metabolism. Trials should specify what toxicities can result and at what measurable levels toxicity should be stopped.</td>
</tr>
</tbody>
</table>

**DISCUSSION**

In this initial offering of the newly designed program, evaluation activities illuminated several issues of potentially broad relevance to simulation-based medical education, related to patient variability, model realism, and the expertise of the facilitator. Accommodating patient variability in an educational simulation, one featured capability of OncoTCap Version 3, is a two-edged sword. It provides important elements of realism but may interfere with learning from comparison. For example, a student who modifies the clinical trial’s regimen and observes a changed clinical trial outcome will be unable to determine how much of the change in outcome is due to chance rather than the study design change that was made. The generation of clinical trial cohorts must be supplemented with the opportunity to try out several ideas on “the same patient” or on “the same cohort of patients.” In the past year, we have designed and implemented OncoTCap Version 4. One of its
new features is to facilitate repeating the same simulation, with a user-chosen design decision, to make the consequences of the design decision discernable despite the effects of random variation. It is challenging to do this without introducing artifacts, but not impossible.6

The degree of realism is another challenge to simulation-based education. Certainly students would rather study real agents rather than apocryphal agents with invented names. However, there is a risk from excessive realism: Students may naively interpret simulation results as actual truths about a specific cancer agent. However, our knowledge of the biology of each agent is not sufficient to make dependable mechanistic predictions across a broad range of possible treatment regimens. If the simulated drugs are labeled with the names of real agents, and if the student is given great freedom in choosing regimens to try, then the simulation may produce results that nature would contradict if the regimen were tested in real life. The potential to mislead students would be serious.

To handle this problem, at the time of case development we presented the course instructor (AB) with several tables of options for qualitative attributes of the biological model. Attributes included options for toxicity profile, pharmacokinetic behavior, mechanism of drug action, anticancer drug dose/response, and drug resistance mechanisms. Table 5 is one example of a set of attributes assigned in a toxicity profile of two anticancer drugs (A and B) used in 2001. The instructor selected the attributes desired for the simulated drugs by placing letters in the boxes at the left. The selections made reflected the instructor’s pedagogic priorities. Thus, well-defined realistic drug characteristics could be modeled, studied, and discussed qualitatively.

In OncoTCap Version 4, currently under development, the software includes a knowledge capture facility that supports the collection and assembly of empirical research results. These assemblages, called validation suites, constitute a list of goodness-of-fit tests used to validate a simulation model. This feature may make it feasible to build workshop exercises around cancer agents from the real world. The subsequent challenge will be to limit the range of student options to the range of model behavior that is validated empirically in the research literature, or to warn the student adequately when that range is exceeded.

Finally, the expertise of the facilitator presents a challenge. The benefit of the presence of a medical oncologist as facilitator is strongly desired by students, but this is expensive and could limit the widespread usage of any educational resource whose goals are not focused on quantifiable "vocational training" skills. The problem may be solvable in the future by intelligent tutoring systems, the technology for which is rapidly developing.7

Instructors interested in evaluating or utilizing OncoTCap for education may find a detailed instructor guide, including download, at <http://www.oncotcap.pitt.edu/tcap>.

References


Table 5. Toxicity Profile Option Table for Hypothetic Anticancer Drugs A and B

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cumulative, permanent (e.g., cardiac, neurologic)</td>
<td>Life threatening but resolves quickly</td>
</tr>
<tr>
<td></td>
<td>Life threatening but fairly easy to manage (e.g., thrombocytopenia)</td>
<td>Driven by concentration peak</td>
</tr>
<tr>
<td></td>
<td>Affected by liver function</td>
<td>Driven by time over threshold</td>
</tr>
<tr>
<td></td>
<td>Affected by renal function</td>
<td>Metabolite drives the toxicity</td>
</tr>
</tbody>
</table>

In OncoTCap Version 4, currently under development, the software includes a knowledge capture facility that supports the collection and assembly of empirical research results. These assemblages, called validation suites, constitute a list of goodness-of-fit tests used to validate a simulation model. This feature may make it feasible to build workshop exercises around cancer agents from the real world. The subsequent challenge will be to limit the range of student options to the range of model behavior that is validated empirically in the research literature, or to warn the student adequately when that range is exceeded.

Finally, the expertise of the facilitator presents a challenge. The benefit of the presence of a medical oncologist as facilitator is strongly desired by students, but this is expensive and could limit the widespread usage of any educational resource whose goals are not focused on quantifiable "vocational training" skills. The problem may be solvable in the future by intelligent tutoring systems, the technology for which is rapidly developing.7

Instructors interested in evaluating or utilizing OncoTCap for education may find a detailed instructor guide, including download, at <http://www.oncotcap.pitt.edu/tcap>.

References