Virtual Reality Skills Training for Health Care Professionals in Alcohol Screening and Brief Intervention

Michael Fleming, MD, MPH, Dale Olsen, PhD, Hilary Stathes, MEd, Laura Boteler, BS, Paul Grossberg, MD, Judie Pfeifer, MEd, Stephanie Schiro, BA, Jane Banning, MSSW, and Susan Skochelak, MD, MPH

Background: Educating physicians and other health care professionals about the identification and treatment of patients who drink more than recommended limits is an ongoing challenge.

Methods: An educational randomized controlled trial was conducted to test the ability of a stand-alone training simulation to improve the clinical skills of health care professionals in alcohol screening and intervention. The "virtual reality simulation" combined video, voice recognition, and nonbranching logic to create an interactive environment that allowed trainees to encounter complex social cues and realistic interpersonal exchanges. The simulation included 707 questions and statements and 1207 simulated patient responses.

Results: A sample of 102 health care professionals (10 physicians; 30 physician assistants or nurse practitioners; 36 medical students; 26 pharmacy, physican assistant, or nurse practitioner students) were randomly assigned to a no training group (n = 51) or a computer-based virtual reality intervention (n = 51). Professionals in both groups had similar pretest standardized patient alcohol screening skill scores: 53.2 (experimental) vs 54.4 (controls), 52.2 vs 53.7 alcohol brief intervention skills, and 42.9 vs 43.5 alcohol referral skills. After repeated practice with the simulation there were significant increases in the scores of the experimental group at 6 months after randomization compared with the control group for the screening (67.7 vs 58.1; $P < .001$) and brief intervention (58.3 vs 51.6; $P < .04$) scenarios.

Conclusions: The technology tested in this trial is the first virtual reality simulation to demonstrate an increase in the alcohol screening and brief intervention skills of health care professionals. (J Am Board Fam Med 2009;22:387–98.)

Training health care professionals to ask or talk with patients about substance use, exposure to interpersonal violence, sexual practices, and other sensitive topics is an ongoing challenge. Despite the critical importance of these areas, several studies have shown that physicians infrequently ask about these topics as part of routine care. For example, in a study of primary care practices, patients with alcohol dependence received the recommended quality of care, including assessment and referral to treatment, only approximately 10% of the time.[1] Fiore et al[2] reported that a population-based survey found that less than 15% of smokers who saw a physician during the last year were offered assistance and only 3% received a follow-up appointment to address tobacco use. In a survey conducted by Elliot et al,[3] 10% of physicians reported screening for domestic violence and only 6% screened all their patients. In a Canadian survey more than 80% of physicians felt they had adequate or excellent medical training in assessing risk behaviors for heart disease and sexually transmitted diseases. The proportion of physicians who felt this way about their training in screening for substance use disorders, family violence, and sexual abuse ranged between 12.7% and 31.6%.[4]

Traditional educational methods used to increase the clinical skills of students and practitioners in these difficult clinical topics include lectures, case-based discussions, evidence-based journal reviews, role plays, videotape playback reviews, e-learning, and standardized patients.[5] Bowman and colleagues[6] and Olsen[7] used simulated patients to improve physician performance on discussing with patients the prevention of sexually transmitted diseases; they were able to demonstrate significant improvements in clinical practice. However, because these methods provide limited ability for learners to repeatedly practice new clinical skills, these teaching methods have significant limitations. Students need the opportunity to repeat the words in multiple clinical situations, to observe patients' reactions to sensitive questions, to practice behavioral intervention statements over and over again, and to receive direct and immediate feedback.

Highly interactive role play simulations have been shown to improve training effectiveness and "boost learning retention rates dramatically."[8,9] Virtual reality simulations offer many potential advantages over traditional educational methods. These advantages include (1) allowing learners to practice the simulation multiple times; (2) ensuring that learners receive a different patient response to the questions and behavioral statements for each virtual reality play; (3) permitting the learner to verbally ask questions and conduct brief intervention through what becomes a "real" clinician-patient interaction using activated voice response; (4) allowing the learner to play the simulation at any time and in any location (portability); (5) scoring of the performance of the learner and providing immediate feedback, both from the patient and a computerized "coach"; (6) a playback feature that can replay the clinician-patient interaction, repeating good interactions or trying purposeful mistakes; (7) providing educational screens that allow the learner to read about basic screening and counseling skills before or during the simulation; and (8) offering the learner course credit or continuing education credits by linking the simulation to an online connection.[10]

The goal of this project was to produce and test a self-contained, off-the-shelf virtual reality simulation system for health care professionals, helping them to improve their clinical skills in the areas of alcohol screening; brief alcohol interventions; and referral for at-risk, problem, and dependent drinkers. This article presents the results of an educational trial designed to test the ability of this system to improve the clinical skills of students and primary care clinicians. A successful demonstration of effect will lead to the development of additional "virtual reality" training simulations for other sensitive behavioral issues such as screening and intervention with underage drinkers; tobacco addiction; illicit drug use; nonprescription opioid abuse; interpersonal violence; sexual risk reduction; and suicide ideation.

Methods

Design

A randomized controlled educational trial was conducted to test the hypothesis of interest. One hundred two health care professionals (students and practitioners) were assigned to the experimental virtual reality simulation program or to a no education control group. The intervention was based on SIMmersion simulation technology (SIMmersion LLC, Columbia, MD).[7,11] Professionals assigned to the intervention group were expected to read the educational materials and to practice the simulation on their personal computer at least 10 times during the 3-month study period. The primary outcomes of interest were changes in the clinical skills of the participants. Clinical skills were assessed using standardized patients. Each participant was tested with 3 different case scenarios at baseline and 6 months after randomization. The case scenarios were developed specifically for this study because there are currently no standard case scenarios that have been tested and validated for alcohol screening, brief intervention, and referral. Our scenarios build on prior work conducted by the principle investigator's (MF) research group.[12]

Participant recruitment

Participants were initially recruited by email and invited to participate in an educational study focused on alcohol screening and brief intervention. Eligible participants included physicians, residents, nurse practitioners, physician assistants, nurse practitioner and physician assistant students, medical students, and pharmacy doctoral students at the University of Wisconsin—Madison. Fourteen percent of the students and clinicians (102 of 731) who were sent a blanket email participated in the trial.

The primary groups were fourth-year medical students and primary care physicians. Participants who responded to the email invitation were contacted by telephone and screened to determine eligibility. These criteria included 18 years of age or older; currently enrolled in a professional training program or currently practicing medicine, nursing, or pharmacy; ability to practice the required number of plays for the simulation; and availability to complete the 6-month posttest at the testing facility in Madison.

Participants who met these criteria were then scheduled to participate in the baseline standardized patient skills test. Written informed consent was obtained at the time of the baseline testing scenario. Randomization also occurred at this time. The study was approved by the University of Wisconsin—Madison Health Sciences Human Subjects Committee. Participants in both groups were paid $50 for completion of the pretest and $50 for the posttest. Intervention participants received an additional $10 for each play up to 10 plays. Eight participants who completed all aspects of the trial were randomly selected to receive an additional $500 for their participation at the end of the trial. The participants in the control group were given the virtual reality simulation program and voice recognition equipment at the end of the 6-month posttest.

Standardized patient testing scenarios and scoring
The testing scenarios were developed by the study research team. Three standardized cases (screening, intervention, and referral) and rating methods were used for the pretest and a separate set of 3 cases were used for the posttests. The scoring items for each of the 6 scenarios described below are listed in Tables 1, 2, and 3.

The pretest screening case discussed a 43-year-old salesman who presented to a new physician for a blood pressure medication refill and evaluation of hypertension. He had a stressful job in a new location, had approximately 4 drinks most nights after work, and occasionally had up to 6 to 8 drinks. He was annoyed by his wife's concern about his drinking and had tried to cut down, but was not successful.

The pretest brief intervention case was of a 48-year-old single, female financial advisor who presented to a new physician for trouble sleeping; she awakened often each night. She had 3 to 4 drinks each night, recently missed a crucial morning business appointment, and sprained her ankle when drinking. She had been previously diagnosed as a problem drinker but did not believe it.

The pretest referral case was of a 48-year-old divorced woman who had been a licensed practical nurse for 30 years. She presented to her physician for a refill of a diazepam prescription to help her relax and sleep. She had 5 to 6 drinks every night, her father and brother were alcoholics, and she had been irritable at work. She did not believe she was an alcoholic and would not go to Alcoholics Anonymous. She was a dependent drinker.

The posttest screening case was of a 40-year-old man who worked as a salesman for a manufacturing company and who presented to his physician to have stitches removed from an injury suffered during a bar fight. He had approximately 4 drinks most nights after work, 6 to 8 on an occasional heavy night, was annoyed by his wife's concern about his drinking and had hangovers several times each month.

The posttest brief intervention case was of a 48-year-old woman with 2 married children and 3 small grandchildren. Her marriage was failing and she did volunteer work for several organizations. She had 3 to 4 drinks most nights and her children were concerned about her drinking and were reluctant to have her babysit her grandchildren on weekends. She had tried to cut down and her father and brother were alcoholics. She was a problem drinker.

The posttest referral case discussed a 47-year-old woman who owned her own business and presented to her physician for increasing fatigue and trouble sleeping. She had 3 to 5 drinks per night and lived with a female partner who is concerned about her drinking. Her father was an alcoholic and she had tried to cut down but was unsuccessful. She was a dependent drinker.

Training of standardized patients
The University of Wisconsin School of Medicine and Public Health Clinical Teaching and Assessment Center (CTAC) was formally established in 1994 and has an active standardized patient (SP) program for research, teaching and clinical assessment. SPs for this project were drawn from a pool of 90 persons who had previously participated as SPs; they were selected based on longevity of experience, interest in the project, time

Table 1. Scores for the Screening Scenario Skills

SIMmersion Items	Control		Experimental		t Statistic	P
	Pretest (mean [SD])	Posttest (mean [SD])	Pretest (mean [SD])	Posttest (mean [SD])		
Alcohol use						
Clinician asked about drinking frequency	4.12 (1.93)	4.17 (1.88)	3.92 (2.08)	5.00 (0.00)	2.900	.005
Clinician asked about drinking amount	4.90 (0.70)	3.75 (2.19)	4.80 (0.98)	4.77 (1.07)	2.770	.007
Clinician asked about 5 drinks in a row	0.98 (2.00)	0.73 (1.78)	0.49 (1.50)	3.14 (2.45)	5.410	<.001
Family history						
Clinician asked about drinking in the family	3.73 (2.20)	4.17 (1.88)	3.92 (2.08)	4.19 (1.87)	0.050	.961
Consequences						
Clinician asked about legal problem due to alcohol	2.06 (2.49)	3.02 (2.47)	1.67 (2.38)	3.72 (2.21)	1.420	.160
Clinician asked about alcohol-related injury	0.39 (1.36)	1.35 (2.25)	0.69 (1.74)	2.79 (2.51)	2.880	.005
Clinician asked about alcohol-related work problems	3.53 (2.30)	2.81 (2.51)	3.53 (2.30)	3.37 (2.37)	1.090	.278
Clinician asked about alcohol-related relationship problems	3.53 (2.30)	4.38 (1.67)	3.14 (2.44)	4.53 (1.47)	0.480	.631
Clinician asked about alcohol treatment history	2.45 (2.52)	2.60 (2.52)	2.84 (2.50)	2.56 (2.53)	0.090	.931
Health issues						
Clinician asked about alcohol-related medical problems	1.08 (2.08)	1.56 (2.34)	1.08 (2.08)	1.16 (2.14)	0.850	.399
Clinician asked about medication	3.33 (2.38)	3.96 (2.05)	3.53 (2.30)	3.37 (2.37)	1.260	.209
CAGE questions						
Clinician asked about attempts to reduce drinking	2.94 (2.49)	2.60 (2.52)	2.45 (2.52)	3.60 (2.27)	1.980	.051
Clinician asked about being annoyed when confronted about drinking	1.67 (2.38)	2.40 (2.52)	1.96 (2.47)	2.33 (2.52)	0.130	.895
Clinician asked about drinking in the morning	2.35 (2.52)	2.29 (2.52)	1.76 (2.41)	3.02 (2.47)	1.400	.166
Communication skills						
Clinician gives reason for alcohol questions	4.41 (1.63)	4.17 (1.88)	4.22 (1.84)	4.07 (1.97)	0.240	.811
Clinician was not sidetracked	4.12 (1.93)	4.90 (0.72)	4.12 (1.93)	5.00 (0.00)	0.950	.347
Clinician was nonjudgmental	4.41 (1.63)	4.38 (1.67)	4.51 (1.50)	4.53 (1.47)	0.480	.631
Overall clinician performance	4.41 (4.08)	4.90 (2.63)	4.61 (3.98)	6.51 (3.37)	2.560	.012

The 17 clinical skills received a score of 5 points if the learner successfully demonstrated skill. The remaining 15 points were based on the learners "overall clinical performance."

availability, and background in a health care-related field. There were 9 used for the pretest and 10 for the posttest. The SPs included 6 women and 4 men. Their ages ranged from 38 to 60 and they had been with the SP program for up to 4 years.

The CTAC director (J.B.) was the primary trainer who taught the SPs to portray the cases and score the clinical skills of the subjects. Each SP participated in two 2-hour training sessions that focused on practicing their assigned role. The CTAC director provided feedback and comments to every SP, and each was expected to portray one case scenario for the pretest and one for the posttest. SPs were also asked to attend several meetings with the researchers to assist in

script development as well as to refine checklist items so they were descriptive, clear, and intuitive to the SPs. The research team created detailed scripts and directions for the checklist completion based on these discussions. The close training with and collaboration between the SPs and members of the research team resulted in an open atmosphere in which SPs' questions could be readily asked and addressed.

A pilot test of the cases, role plays, checklists, technology, and logistics took place in September 2007 with 3 participants and 8 SPs either portraying roles or viewing the sessions. None of the 3 participants involved in the pilot participated in the trial. Further modification of the scripts and scoring items were facilitated by immediate after-pilot

Table 2. Scores for the Brief Intervention Scenario Skills

	Control		Experimental			
SIMmersion Items	Pretest (mean [SD])	Posttest (mean [SD])	Pretest (mean [SD])	Posttest (mean [SD])	t Statistic	P
Advise						
Clinician is concerned about drinking	4.02 (2.00)	4.69 (1.22)	4.41 (1.63)	4.65 (1.29)	0.140	.891
Clinician offered rationale for concern	4.61 (1.36)	4.27 (1.78)	4.41 (1.63)	4.53 (1.47)	0.770	.446
Clinician advised to reduce drinking	4.71 (1.19)	4.79 (1.01)	4.90 (0.70)	4.77 (1.07)	0.110	.912
Assessment skills						
Clinician asked about drinking pros	1.76 (2.41)	1.88 (2.45)	1.47 (2.30)	2.79 (2.51)	1.760	.082
Clinician asked about drinking cons	0.20 (0.98)	0.63 (1.67)	0.49 (1.50)	2.79 (2.51)	4.890	<.001
Clinician asked about readiness to change	4.31 (1.74)	3.65 (2.25)	4.22 (1.84)	4.88 (0.76)	3.440	<.001
Clinician asked about alcohol harms	0.10 (0.70)	0.42 (1.40)	0.29 (1.19)	0.81 (1.87)	1.160	.251
Clinician asked about stress and drinking	1.67 (2.38)	0.42 (1.40)	1.08 (2.08)	0.58 (1.62)	0.520	.604
Negotiation skils						
Clinician asked to cut down to specific amount of alcohol	3.63 (2.25)	1.88 (2.45)	2.84 (2.50)	2.09 (2.50)	0.420	.675
Clinician asked about follow-up visit	3.53 (2.30)	2.71 (2.52)	3.63 (2.25)	3.60 (2.27)	1.780	.079
Clinician asked about social support	0.98 (2.00)	1.04 (2.05)	0.98 (2.00)	1.28 (2.21)	0.530	.596
Clinician offered continued support	3.43 (2.34)	2.60 (2.52)	3.73 (2.20)	2.56 (2.53)	0.090	.931
Clinician offered education material	1.47 (2.30)	0.94 (1.97)	0.98 (2.00)	0.58 (1.62)	0.930	.353
Communication skills						
Clinician used open-ended questions	3.24 (2.41)	3.65 (2.25)	3.33 (2.38)	3.95 (2.06)	0.680	.499
Clinician was nonjudgmental	4.12 (1.93)	4.27 (1.78)	3.82 (2.14)	4.53 (1.47)	0.770	.446
Clinician was empathic	4.31 (1.74)	4.17 (1.88)	4.02 (2.00)	4.42 (1.62)	0.680	.498
Clinician dealt with resistance	3.43 (2.34)	3.75 (2.19)	3.43 (2.34)	3.26 (2.41)	1.030	.308
Overall clinician performance	4.22 (3.37)	5.94 (3.95)	4.51 (3.50)	6.28 (3.79)	0.420	.676

The 17 clinical skills listed above received a score of 5 points if the learner successfully demonstrated skill. The remaining 15 points were based on the learners "overall clinical performance."

discussion and videotape review with the SPs and the research team observing the pilot and subjects.

Testing procedure

There were 10 sessions used to conduct the pretest standardized patient scenario during a 6-week period in November and December of 2007, with 8 to 12 participants tested during each session. Each of the 102 participants interviewed 3 SPs. The participants were given 5 minutes to read an abbreviated medical record on the outside of the door before entering the examination room and interviewing the SP. Each research participant had 15 minutes to interview and/or counsel the SP. The SP had 5 minutes to score the checklist before the next clinician entered the room. Research participants completed the pretest in 60 minutes. The posttest sessions occurred during April and May of 2008 using similar testing methodology. All subject-SP sessions were videotaped, and a 20% sample of the video taped sessions was reviewed by a panel of

research team members to determine the validity of the SP scoring. The reviews found strong agreement between the ratings of the SPs and research team. The degree of agreement (Kappa) between the SP and research team panel's 20% review was 0.95.

Virtual reality simulation

The virtual reality simulation software used for this study was developed by Dr. Dale Olsen at the Johns Hopkins University Applied Physics Laboratory. The stand-alone software modified for this study consisted of 6 elements. The first element was an educational program that could be read by the participant to give basic background on alcohol screening and intervention. This was based on the National Institute on Alcohol Abuse and Alcoholism's 2005 clinician's guide and consisted of 20 screens of text. The second element consisted of 707 questions and statements learners were able to ask the simulated patient to conduct counseling or make a referral to a treatment center. The third

Table 3. Scores for the Referral Scenario Skills

SIMmersion Items	Control		Experimental		*t* Statistic	*P*
	Pretest (mean [SD])	Posttest (mean [SD])	Pretest (mean [SD])	Posttest (mean [SD])		
Advise						
Clinician advised to reduce drinking	4.22 (1.84)	4.69 (1.22)	3.73 (2.20)	5.00 (0.00)	1.670	.098
Clinician offered reasons for concern	3.82 (2.14)	4.90 (0.72)	3.92 (2.08)	4.65 (1.29)	1.130	.261
Clinician discussed consequences	2.16 (2.50)	3.85 (2.12)	2.06 (2.49)	3.37 (2.37)	1.020	.309
Assessment skills						
Clinician asked about treatment history	1.47 (2.30)	0.83 (1.88)	1.27 (2.20)	0.81 (1.87)	0.050	.961
Clinician asked about readiness to change	2.06 (2.49)	3.85 (2.12)	1.27 (2.20)	4.19 (1.87)	0.790	.433
Negotiation skills						
Clinician asked about joining AA group	1.76 (2.41)	3.23 (2.42)	2.25 (2.51)	2.21 (2.51)	1.970	.052
Clinician asked about insurance	0.20 (0.98)	0.31 (1.22)	0.20 (0.98)	0.23 (1.07)	0.330	.742
Clinician asked about willingness to receive alcohol assessment	2.45 (2.52)	4.69 (1.22)	3.04 (2.47)	4.30 (1.75)	1.230	.224
Clinician scheduled alcohol assessment	1.27 (2.20)	1.25 (2.19)	1.76 (2.41)	1.98 (2.47)	1.490	.141
Clinician asked to come for follow-up	1.76 (2.41)	2.40 (2.52)	1.76 (2.41)	3.37 (2.37)	1.900	.061
Clinician asked about social support	1.18 (2.14)	1.67 (2.38)	1.37 (2.25)	2.09 (2.50)	0.830	.407
Clinician offered continued support	1.08 (2.08)	3.44 (2.34)	1.67 (2.38)	3.72 (2.21)	0.590	.555
Clinician tried to match referrals to preferences	1.47 (2.30)	4.48 (1.54)	1.37 (2.25)	4.30 (1.75)	0.510	.613
Communication Skills						
Clinician used open-ended questions	2.75 (2.51)	4.27 (1.78)	2.16 (2.50)	4.53 (1.47)	0.770	.446
Clinician responded positively to resistance	3.53 (2.30)	4.90 (0.72)	3.53 (2.30)	4.77 (1.07)	0.680	.499
Clinician was nonjudgmental	4.41 (1.63)	4.79 (1.01)	4.31 (1.74)	4.77 (1.07)	0.110	.912
Clinician was empathic	3.82 (2.14)	4.38 (1.67)	3.73 (2.20)	4.65 (1.29)	0.880	.384
Overall clinician performance	4.12 (3.56)	6.56 (3.29)	3.53 (3.91)	7.09 (3.82)	0.710	.478

The 17 clinical skills listed above received a score of 5 points if the learner successfully demonstrated skill. The remaining 15 points were based on the learners "overall clinical performance."
AA, Alcoholics Anonymous.

element was a set of 1207 responses developed for the simulated patient to respond to the learner (an actress video recorded these responses with varied mood and affect in a production studio during a 5-day period). An algorithm embedded in the program was programmed to respond based on the type and appropriateness of the question or statement made by the learner, as well as the history of the conversation between the character and the learner. Negative or inappropriate questions or statements were linked to audio and visual anger or mood changes in the SP.

The fourth element was the on-screen help agent. The agent is an action figure in the corner of the screen who intermittently displays hand and body signals to indicate especially good or not-so-good questions posed by the learner. The fifth element was an instant replay feature. The sixth element was scoring and feedback to the learner about their performance during the play. The basic computer screen for the simulation is presented in Figure 1. For this simulation the program was designed so that 40% of the time the character in the simulation would be an at-risk drinker, 40% of the time they would be a problem or dependent drinker, and 20% of the time they would be a low-risk drinker (based on criteria developed by the National Institute on Alcohol Abuse and Alcoholism).

Participants were asked to conduct face-to-face alcohol screenings, brief interventions, and referrals with the simulated character using a microphone or a computer mouse to communicate. The questions and statements were scripted to include a variety of natural choices. For each scripted question or statement there were multiple simulated character responses available. The simulated character's brain selects a response based on the level of rapport developed by the physician, the character's risk level, previously discussed information, and
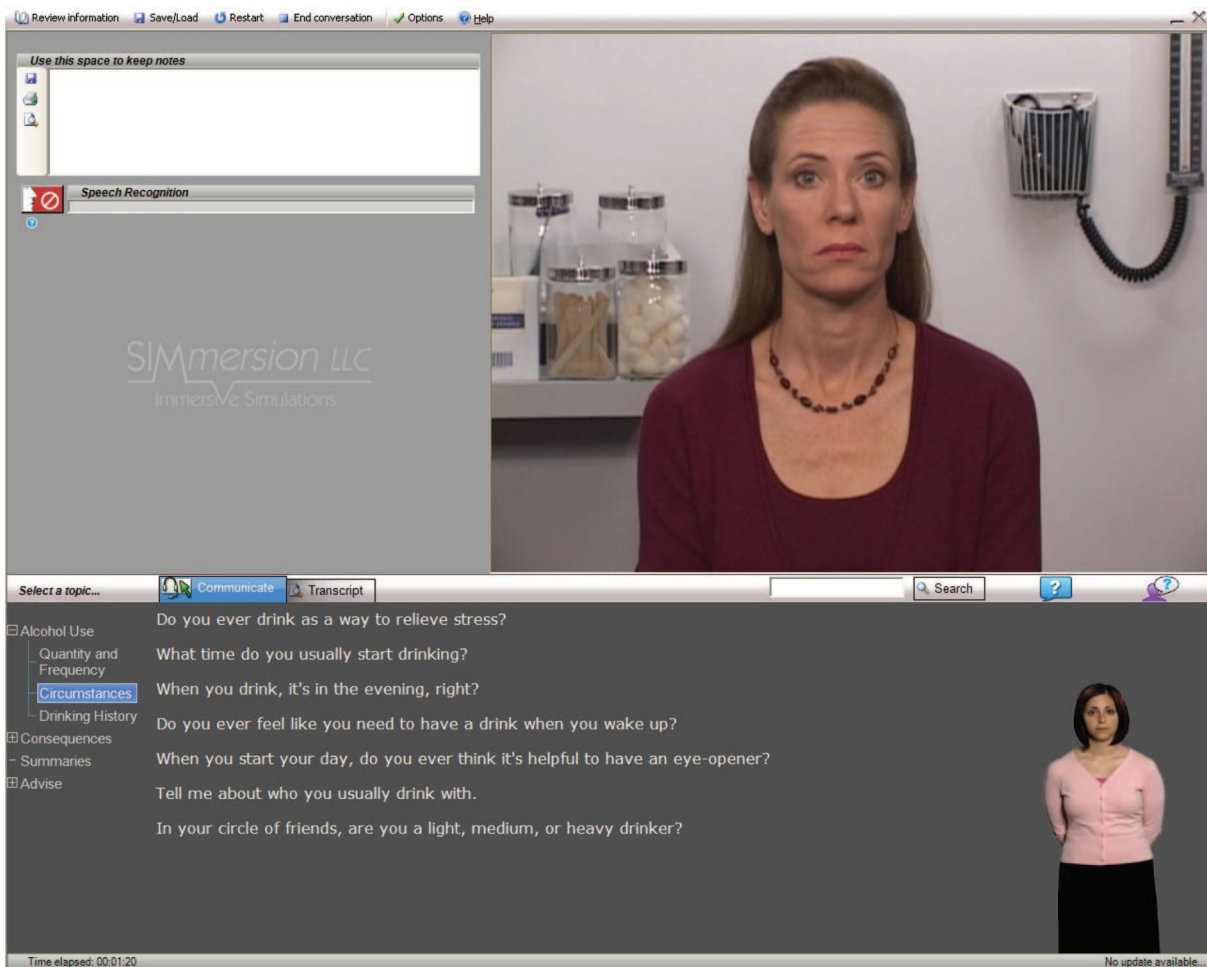
Figure 1. The basic computer screen for the simulation.

chance. For example, if the physician selects a series of inappropriate options, the simulated character will become curt and uncooperative; if, however, the physician selects a series of appropriate options, the character will become friendly and forthcoming. This realistic emotional variation allows the simulated character to emulate actual human behavior.

Participants received feedback from an on-screen help agent who provided nonverbal cues regarding the user's choice of questions. In addition, the participant could click help buttons for assistance with question choices and character responses. The system scored the participant's performance and an instant replay feature enabled users to review portions of dialogue or their entire conversation. At the end of the conversation, the participant would have to decide which type of drinker (randomly selected by the program) they were talking to and were scored on their accuracy.

Each participant was expected to screen the patient for alcohol use and decide if the patient required a brief intervention and a referral for treatment. Because each of the 3 patients would require varying amounts of time, a completed play was based on the number of statements and questions that were used by the participant, with 10 statements set as the minimum requirement for one play. The guidelines could be accessed at anytime during the play as a reference for the participant.

Plays were tracked using SIMmersion's (SIMmersion, LLC) online tracking system, with the goal being that each experimental group participant would play at least 10 times in the 3 to 4 months before the final posttest. After pilot testing by the expert panel there was a general consensus that 10 plays was a minimum number of plays needed to take advantage of the various patient scenarios and responses built into the simulation. The experimental group participants were contacted by re-

Table 4. Sociodemographic Characteristics of Study Population

Characteristic	Control Group (n = 51)	Simulation Group (n = 51)	Total (n = 102)
Age* (years)			
20–30	27 (53)	32 (63)	59 (58)
31–40	15 (29)	8 (16)	23 (23)
>40	9 (18)	11 (22)	20 (20)
Sex*			
Male	16 (31)	10 (20)	26 (26)
Female	35 (69)	41 (80)	76 (75)
Student status*			
Medical student	18 (35)	18 (35)	36 (35)
Nurse practitioner, physician assistant, or pharmacy student	13 (26)	13 (26)	26 (26)
Practicing clinician*			
Physician	5 (10)	5 (10)	10 (10)
Nurse practitioner or physician assistant	15 (29)	15 (29)	30 (29)
Previous alcohol intervention training* (hours)			
0	44 (86)	51 (100)	95 (93)
≤5	6 (12)	0 (0)	6 (6)
>10	1 (2)	0 (0)	1 (1)
Number of plays with simulation (minimum 20 min per play)			
0	-	4 (8)	-
1–4	-	7 (14)	-
5–9	-	10 (20)	-
≥10	-	30 (59)	-

*There were no significant differences ($P < .05$) between groups.
All data provided as n (%).

search staff once during the practice period to ensure they were using the simulation software correctly and to aid in solving any issues.

The questions and SP responses were written and developed by an expert panel at the University of Wisconsin. The expert panel consisted of 14 University of Wisconsin—Madison primary care clinicians, addiction medicine physicians, psychologists, and members of the SIMmersion team with significant expertise in substance abuse. Members of the panel also played and pilot-tested a number of versions of the virtual reality simulation as it was developed. The simulation took approximately 9 months to develop before testing.

Statistical analysis

The demographic variables for the experimental and control groups were described by way of frequencies (%). Univariate analysis was used to assess potential differences on sex, age, clinician versus student status, and prior alcohol training. As noted in Table 4, there were no significant differences between groups on these 4 variables.

The primary outcomes were the intervention, screening, and referral scores from the SP scenarios. As noted in Tables 1, 2, and 3, the clinical skills of the participants were assessed using a set of clinical criteria developed for the screening, brief intervention, and referral scenarios. Eighty-five points was allocated to 17 specific skills criteria, with 5 points for each. The SPs were instructed to score the skill done as a simple yes (the learner demonstrated the skill) or, alternatively, no (they did not demonstrate the skill). Participants received 5 points or 0 points for each skill. No partial credit scores were given. Fifteen points were used to rate the clinicians overall performance. These scores (0 to 100) were aggregated to create a total score for each scenario and were described with means and standard deviations.

Ninety-one of 102 participants completed the 6-months posttest. The primary reasons the 11

Table 5. Scores for the Screening, Brief Intervention, and Referral Testing Scenarios

SIMmersion Items	Control Group (n = 51)		Simulation Group (n = 51)		t Statistic	P
	Pretest (mean [SD])	Posttest (mean [SD])	Pretest (mean [SD])	Posttest (mean [SD])		
Screening						
Total score	54.41 (15.58)	58.13 (15.51)	53.24 (16.09)	67.67 (12.42)	3.43	<.001
Total score change	-	3.71 (19.16)	-	14.44 (16.17)	3.06	.003
Intervention						
Total score	53.73 (19.34)	51.67 (18.09)	52.55 (13.86)	58.37 (15.89)	1.99	.050
Total score change	-	2.06 (23.27)	-	5.82 (20.13)	1.83	.070
Referral						
Total score	43.53 (17.79)	64.48 (12.01)	42.94 (14.28)	66.05 (14.02)	0.61	.546
Total score change	-	20.95 (17.34)	-	23.11 (20.44)	0.57	.567

The Total potential score for each scenario was from 0 to 100. Intention to treat procedures were followed and baseline values were imputed for the 11 professionals who did not participate in the 3 posttest scenarios.

participants did not complete the posttest included scheduling conflicts with patient care and course work, relocation, and illness. Intention to treat analysis was followed. Baseline scores were imputed to the missing data in the posttest scores for the 11 participants who did not complete it. We elected to use the most conservative method to handle missing follow-up data. All 102 participants originally randomized into the trial were included in the outcome analysis.

The mean values for experimental and control groups on the posttest were compared with t tests to derive effect size for the educational trial on the 3 scenarios. The t tests were executed separately for all items on the intervention, referral, and screening scales. The sample was too small to assess a dose-response affect between the number of plays and changes in clinical behavior. We stratified the results on each of covariates in Table 4 and found no statistical association between these variables and the primary outcomes. The analyses were performed with SAS software (version 9.1 for Linux; SAS Institute, Inc., Cary, NC).

Results

Table 4 provides a general description of the 102 trial participants. As noted previously, the majority of the research participants were medical students and other health care professional students. Ten primary care physicians (4 family physicians, 4 internal medicine physicians, and 2 pediatricians) participated in addition to 30 nurse practitioners and physician assistants. As in most health care professions, the majority of participants were women. Only 5% reported previous training in alcohol screening and brief intervention.

Table 5 presents the primary results of the trial. The scoring was set up to range between 0 and 100 points for each case. Baseline scores were similar for each group across all 3 scenarios. The posttest scores demonstrated significant differences between groups for alcohol screening ($P < .001$) and brief intervention skills ($P < .04$). The screening skills in the control group increased by 3.7% whereas the skills in the experimental group improved by 14.4% during the 6-month period. The number of participants who inquired about the frequency of alcohol use increased from 40 to 51 in the experimental group, with no change in the control group (42 to 43). The brief intervention skills went down in the control group by 2.1% and increased 5.7% in the experimental group. Although there were significant changes in referral skills on the pre- and posttests, there were no differences between groups.

Tables 1 through 3 present the score and differences for the individual items contained in each of the scenarios. Table 1 illustrates each individual screening item on which participants were scored. The items with significant changes between pre- and posttest included asking about quantity of alcohol use, the frequency of alcohol use, and the frequency of heavy drinking. Participants also asked more often about alcohol-related injuries and

prior treatment. Table 2 lists the items scored for the brief intervention scenario. Items with significant changes focused on drinking cons and readiness to change. As shown in Table 3, there were some minor differences between the control and intervention groups in the frequency of referral to Alcoholics Anonymous and making a follow-up appointment.

Discussion

The technology tested here is in a true sense a "virtual reality teaching method" that allows a learner to engage in a "real doctor patient interaction." The algorithm on which the technology is based allows for an unlimited number of variations in the interaction so that every time the learner plays the program the situation is novel and different. Because of the large number of potential responses to questions and counseling statements, the patient can respond in a nearly limitless number of ways. There is also a realistic emotional variation in the patient response that allows the patient to react to the type and sequence of questions and statements. These emotional variations include negative or positive facial expressions, nonverbal body postures, negative vocalizations, changes in eye contact and expressions of affect that change the atmosphere of the interaction. The simulation attempted to overcome concerns about the "reality" of patient simulations,[13] allowing learners to "practice" on fake patients before applying these new skills to real patients where patient safety is a concern.[14]

The 14 members of the expert panel who practiced and played the program reported that the interview felt like entering an office examination room and dealing with a real patient. Similar unsolicited comments were made by the students and physicians in the intervention group. The program made learning fun and interactive, especially for a topic that is difficult and sensitive. For a topic like substance abuse, where learners often roll their eyes and become somnolent when being taught, this technology represents a refreshing, innovative way to teach and enhance clinical skills for a variety of health care professionals.

The most robust finding of the study was an improvement in alcohol screening skills. This finding may be related to nature of the simulation that began with screening and then moved into brief intervention and referral. With a limited number of plays (minimum of 10), learners may not spend sufficient time practicing brief intervention and referral skills. Although we were not able to test the dose effect of additional plays, future research may want to focus on the brief intervention and referral skills training portions of the simulation.

How can this virtual reality program be used to improve the knowledge and skills of health care professionals? First, the program could function as part of a core curriculum for teaching medical, physician assistant, nursing, and pharmacy students about how to conduct alcohol screening and intervention. Other parts of the curriculum could include case discussions, e-learning sites, an evidenced-based review of clinical protocols, and SP testing. The program could serve as the platform for a more comprehensive program that could include rotations on consult services, use of these skills on clinical rotations with feedback by faculty supervisors, supervised assessments in addiction medicine programs, as well as opportunities to practice these new skills in community-based prevention and treatment programs.

Second, because the program is designed as a stand-alone resource, the protocol could be used for continuing medical education (CME) for practicing physicians. In a review of studies on formal CME programs, Davis et al[15] presented evidence that interactive CME sessions that enhance participant activity and provide the opportunity to practice skills can effect change in professional practice and, on occasion, health care outcomes. Although the program is easy to install and learn how to use, a brief orientation by local information technology staff may be helpful in overcoming challenges many clinicians have with new computer technology. Third, for health care systems or specific groups such as hospital-based trauma surgeons, the program could be used to meet the requirements for national certifying organizations like the Joint Commission or American Trauma Society.

The strengths of this study include the random assignment of a control group, a large diverse sample of learners, state-of-the-art measurement of clinical skills, intention-to-treat procedures, and collection of 90% of posttest follow-up information. Weakness of the study included challenges in measuring changes in clinical behavior skills. There are no standard methods to measure changes in alcohol screening and intervention skills, and we

had to develop our own SP scenarios and scoring methods. Although we pilot tested the scenarios and used the research team to review 20% of the tapes to ensure consistency, scoring remained a challenge. The screening and brief intervention scenarios seemed to work well with large positive outcomes in the intervention group. The large change in the referral scenarios for both groups suggests a problem with the posttest scenario that needs further assessment.

Another potential limitation is the generalizability of a volunteer sample compared with a general sample of learners. On the one hand, one could argue that paid volunteers are more motivated or are more likely to change their clinical skills. However, in this case we make the opposite argument. A student or clinician who is required to play the simulation is more likely to learn the material and get a higher score on the posttest than a volunteer. Our volunteers did not receive a grade or feel any pressure to pass the posttest or to learn how to screen and conduct brief intervention. Grades and a passing requirement are a powerful motivator to perform. Another argument is the observation that the pretest baseline standardized patient scores on alcohol intervention skills are likely to be lower in a general sample than in a volunteer sample. Most of our volunteers were interested in leaning more about alcohol intervention and had fairly high baseline pretest scores. This created a ceiling effect that limited our ability to demonstrate change. Based on these arguments, we expect clinicians and students who are required to play the simulation and pass a certain level of proficiency would be likely to improve their skills at least as much as our volunteer sample.

Conclusion

The technology successfully tested in this study offers great promise. Virtual-reality, game-based teaching methods are ideally suited to increasing the behavioral skills of health care clinicians. It is difficult to ask patients about personal issues (substance use, sexual practices, violence, depression) when they made an appointment for a blood pressure check or a headache. Although patients appreciate concern and caring from their clinicians they also expect personal questions to be asked with skill, empathy, and confidentiality. Clinicians who do not have the skills to inquire or talk about these sensitive topics often generate fear and resistance on the part of the patient. Patients' reactions to being asked about these topics can be strong and in some cases carry the risk of harm. Virtual-reality simulation offers learners the opportunity to practice and develop skills before trying to apply these skills with real patients.

References

1. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003;348:2635–45.
2. Fiore MC, Bailey WC, Cohen SJ, et al. Treating tobacco use and dependence. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services, Public Health Service; June 2000.
3. Elliott L, Nerney M, Jones T, Friedmann PD. Barriers to screening for domestic violence. J Gen Intern Med 2002;17:112–6.
4. Maheux B, Haley N, Rivard M, Gervais A. Do physicians assess lifestyle health risks during general medical examinations? A survey of general practitioners and obstetrician-gynecologists in Quebec. CMAJ 1999;160:1830–4.
5. Nardone DA, Schiner CL, Guyer-Kelley P, Kositch L. Use of computer simulations to teach history taking to first year medical students. J Med Educ 1987;62:191–3.
6. Bowman MA, Russell NK, Boekeloo BO, Rafi IZ, Rabin DL. The effect of educational preparation on physician performance with a sexually transmitted disease-simulated patient. Arch Intern Med 1992; 152:1823–8.
7. Olsen DE. Interview and interrogation training using a computer-simulated subject. Presented at the 19th Interservice/Industrial Training, Simulation, and Education Conference (I/ITSEC), December 1–4, 1997, Orlando, Florida.
8. Bohle S. Simulations: the next generation of e-learning. Training 2005;42(1):1–6. Available at http://marshwhiteinc.com/Documents/Simulations%20the

%20Next%20Generation%20of%20Learning.pdf. Accessed July 8, 2008.

9. Kneebone R, Nestel D, Wetzel D, et al. The human face of simulation: patient-focused simulation training. Acad Med 2006;81:919–24.

10. Triola M, Feldman H, Kalet AL, et al. A randomized trial of teaching clinical skill using virtual and live standardized patients. J Gen Intern Med 2006;21:424–9.

11. Olsen DE, Phillips RG, Sellers WA. The simulation of a human subject for interpersonal skill training. Presented at Interservice/Industrial Training, Simulation and Education Conference (I/ITSEC), November 27 through 30, 2000, Orlando, Florida.

12. Kokotailo PK, Langhough R, Neary E, Matson SC, Fleming MF. Improving pediatric residents alcohol and other drug use clinical skills, an experimental curriculum. Pediatrics 1995;96:99–107.

13. Hanna M, Finns J. Power and communication: why simulation training ought to be complemented by experiential and humanist learning. Acad Med 2006; 81:265–70.

14. Ziv A, Wolpe PR, Small SD, Shimon G. Simulation-based medical education: an ethical imperative. Acad Med 2003;78:783–8.

15. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal Continuing Medical Education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? JAMA 1999;282:867–74.