Medical Informatics and the Internet in Medicine
March 2006; 31(1): 1 – 8

Live or computerized simulation of clinical encounters: Do clinicians work up patient cases differently?

MATHIEU R. NENDAZ[1,2], BELEN PONTE[1], ANNE M. GUT[1], ARNAUD PERRIER[1,2], MARTINE LOUIS-SIMONET[1], ALAIN F. JUNOD[1] & NU V. VU[2]

[1]*Clinique de Médecine 1, University Hospitals, Geneva, Switzerland* [2]*Unit of Development and Research in Medical Education, University of Geneva Faculty of Medicine, Geneva, Switzerland*

(Received July 2004; accepted June 2005)

Abstract
Computer simulation of clinical encounters is increasingly used in clinical settings to train patient work-up. The aim of this prospective, controlled study was to compare the characteristics of data collection and diagnostic exploration of physicians working up cases with a standardized patient and in a computerized simulation. Six clinicians of different clinical experience in internal medicine worked up three cases with a standardized patient and through a computer simulation allowing free inquiry. After each encounter, we asked the subjects to justify the information collected and to comment on their working diagnoses. The characteristics of data collected and working diagnoses generated were assessed and compared, according to the simulation method used. In the computer simulation, physicians limited their data collection and focused earlier and more specifically on information and working diagnoses with high levels of relevance. They reached a similar diagnostic accuracy and made decisions of a similar relevance. Computer simulation with a free-inquiry approach reproduces the data collection and the diagnostic exploration observed in a standardized-patient simulation and promotes an early collection of relevant data. Its contribution to extend the competence of learners in clinical settings should be further evaluated.

Keywords: *Live simulation, computerized simulation, clinical encounters, clinicians, patient cases, clinical reasoning*

1. Introduction

Computer virtual-reality simulation has enjoyed a growing importance in medical education, particularly for high-risk technical activities (e.g. surgery or endoscopic procedures, anaesthesiology, emergency [1 – 4]), because it offers clinicians a safe practice for difficult tasks. As faulty clinical data collection may lead to wrong diagnoses or decisions [5] with potential damage to patients, simulation of clinical reasoning has also been of interest to medical educators. It is often based on standardized patients (SP), with live actors performing the role of the patient, but computer programs are increasingly used to complete learners'

Correspondence: M. R. Nendaz, University Hospitals, Department of Internal Medicine, 1211 Geneva 14, Switzerland. Tel: +41-22-372-9102. Fax: +41-22-372-9116. E-mail: mathieu.nendaz@hcuge.ch

ISSN 1463-9238 print/ISSN 1464-5238 online © 2006 Taylor & Francis
DOI: 10.1080/14639230500236701

experiences [6]. However, the extent to which a computerized simulation permits users to reproduce the collection of clinical data and the exploration of diagnostic hypotheses observed in a live situation needs to be assessed [4]. We conducted a prospective, controlled exploratory study to compare the characteristics of data collection and diagnostic exploration of physicians working up the same set of patient cases presented, first in a live simulation with a SP, then in a computerized simulation.

2. Subjects and methods

The Virtual Internet Patient Simulation (VIPS®, www.swissvips.ch) was the computerized simulator chosen for our study because it allows cases to be constructed in a way fostering a free inquiry, a condition shown to better reflect clinical practice [7]. The software opens with the picture, a short bio-sketch, and the chief complaint of the patient on the upper-left part of the screen. On the lower-left part of the screen, the user types the questions in a specific frame. The right part of the screen is devoted to the list of the questions asked, along with their answers. For each request made by the physician, the software performs free-text recognition of key words and gives the answer, based on a predefined database developed by the case author. If there is ambiguity about the question asked, the software proposes one or several options among which the user must confirm a choice. For the physical examination, the user selects the localization of interest of the human body on a patient outline and then selects the instrument to use (e.g. hands, stethoscope . . .) and the manoeuvre to be performed (e.g. palpation, auscultation . . .).

We created three case scripts, based on real patient charts, by predetermining the answer to each specific question. The chief complaints were chronic diarrhoea, headache, and weight loss. Each answer could represent useful information (e.g. abdominal pain in the diarrhoea case), neutral information (e.g. good general health), or distracters (e.g. menopause symptoms in the headache case).

The inquiry process in both the SP and computer simulations involved the collection of information pertaining to history and physical examination, and diagnostic or therapeutic decisions. Three internists (with 15–34 years of experience) and three second-year internal medicine residents first worked up three cases presented in a SP simulation, in which a person played the role of the patient and gave the answers to the physician, according to the predetermined case script. At least 3 months later (median 18, range 3–33), the same individuals worked up the same cases using the VIPS. In addition to the elapsed time between both simulations, we minimized the recall bias by modifying the elements of the cases that did not influence the patient's work-up (e.g. biographical and non-relevant medical information [8]). Moreover, the participants received no feedback after the encounters with the SP and were not aware that similar cases could be used with the computer simulation.

A record of each encounter was kept, on an audiotape for the SP simulation and on an electronic file for the computer simulation (Appendix). After each encounter, the physicians immediately listened to the audiotape or immediately reviewed the electronic logging and were asked by an investigator to indicate the purposes justifying the collection of each piece of information and their underlying working diagnoses.

Two sets of analyses were conducted: first, a comparison between the two simulation formats of the clinicians' diagnostic accuracy and the relevance and the purpose of the information collected, the working diagnoses generated, and the therapeutic decisions made; second, an analysis of the extent to which items generated with the SP were reproduced in the VIPS. We classified the purposes stated by the participants for each piece of information collected into four categories: routine checking for the presence of symptoms or signs, clarifying the patient's chief complaint or symptoms, testing specific diagnostic hypotheses, and other purposes. The diagnostic accuracy and relevance of the items were evaluated by using a published method [9,10] based on the degree of concurrence among six expert internists (with 15–34 years of experience) who previously worked up the same cases with a SP. Items were categorized into one of the following groups: those generated by all experts of the reference panel (highly relevant, 100% concurrency), those generated by 99% to 80%, those generated by 79% to 40%, those generated by less than 40%, and those not generated by any physician of the expert panel. The characteristics of the clinicians' encounters using each method (SP versus VIPS) were compared by paired t-tests or Wilcoxon signed ranks tests for continuous variables, depending on their distribution, and McNemar or chi-square tests for categorical variables.

3. Results

3.1. Characteristics of data collected, working diagnoses generated, and decisions made

The mean duration of each encounter was similar in both simulation methods. Clinicians using the computer simulation elicited, by encounter, a significantly lower number of history and physical examination items and working diagnostic hypotheses. In contrast, there was no difference in the number of decisions made (Table I). Participants collected a larger number of distinct, highly relevant information items with the SP than with the VIPS: 10.3 (95% CI 9.1–11.5) versus 8.3 (95% CI 7.2–9.4, $p = 0.003$), corresponding, respectively, to 79% and 63% of the maximal number of highly relevant items achieved by the expert reference group. However, the proportion of highly relevant data was significantly larger in the VIPS than in the SP format (Table II). The repartition of the 926 pieces of history information collected with the SP among the different purposes of data collection was: routine checking for the presence of symptoms or signs, 17%; clarifying the patient's chief complaint or symptoms,

Table I. Characteristics, by encounter, of the simulation method used: Standardized Patient (SP) versus the Virtual Internet Patient Simulation (VIPS).

	SP		VIPS		
	Mean	95% CI	Mean	95% CI	p[a]
Encounter duration (min)	17.1	15.1–19	18.1	14.3–22.0	0.50
Total number of information items collected	75.6	67.1–84.8	47.4	37.9–56.8	<0.0001
Total number of working hypotheses generated	15.1	12.6–17.6	8.1	5.7–10.1	<0.0001
Total number of diagnostic or therapeutic decisions made	7.4	4.7–9.9	9.1	6.9–11.4	0.20
Number of distinct highly relevant data collected (history and physical examination)[b]	10.3	9.1–11.5	8.3	7.2–9.4	0.003
Number of highly relevant working diagnoses generated[b]	2.9	2.1–3.8	2.6	1.9–3.2	0.31
Number of highly relevant diagnostic or therapeutic decisions made[b]	2.0	1.3–2.7	2.3	1.5–3.1	0.44

[a]Paired t-tests or Wilcoxon signed ranks tests.

[b]Highly relevant items are those elicited by all members of a reference group of experts in internal medicine who previously solved the same cases with a standardized patient.

Table II. Number and proportions of items, according to their levels of relevance and the simulation method.

Level of relevance (% concurrence)[a]	SP ($n = 1361$)		VIPS ($n = 693$)	
	n	%	n	%
0	176	13	90	13
1–39	347	26	146	21
40–79	354	26	170	25
80–99	181	13	93	13
100	303	22	194	28

[a]Level of relevance based on the degree of concurrence among expert internists of a reference group who previously worked up the same cases with an SP: highly relevant items are those elicited by all members of this reference group. SP: Standardized Patient Simulation; VIPS: Virtual Internet Patient Simulation. $\chi^2(4) = 10.4$, $p = 0.03$.

31%; testing specific diagnostic hypotheses, 50%; and other purposes, 2%. For the 365 pieces of history information collected with the VIPS, the repartition was, respectively, for each purpose: 8%, 33%, 58%, and 1%. These repartitions were significantly different, according to the simulation method ($p < 0.0001$). For physical examination items and decisions, the distribution among purposes was no more significantly different between both methods.

There was no difference between the two methods in the number of relevant working diagnoses generated and relevant management decisions made (Table I). The proportion of highly relevant working diagnoses was also significantly larger in the VIPS than in the SP format: 20% of the 268 diagnostic hypotheses generated with the SP and 32% of the 146 diagnostic hypotheses generated with the VIPS were highly relevant ($p < 0.0001$). Diagnostic accuracy was similar in both simulation methods: 13 of 18 final diagnoses were correct in the SP simulation (72%, 95% CI 49–86), versus 11 of 18 in the computer simulation (61%, 95% CI 39–79, $p = 0.45$).

Highly relevant history information was collected earlier during the VIPS work-up: after a median of 10 items of information collected (interquartile range 10), versus 20 with the SP (interquartile range 21, $p < 0.0001$). Similarly, the correct final diagnosis was generated after a median of five questions (interquartile range 5) in VIPS, compared with 19 with the SP simulation (interquartile range 46, $p = 0.004$).

3.2. Reproduction of data collection and diagnostic hypothesis exploration

Did clinicians collect different data in both simulations? The distribution of the 1468 total distinct items (history, physical examination, and decisions) generated by either format is displayed in Table III. Among the 1204 distinct items generated in the SP simulation, 416 (35%, 95% CI 32–37) were reproduced in the VIPS. While this proportion was low for poorly relevant items, it increased linearly with their levels of relevance to reach 69% (95% CI 63–75) for highly relevant information ($p < 0.0001$, Figure 1). For the working diagnoses ($n = 268$), the overall percentage reproduced in VIPS was 34% (95% CI 28–39), but reached 70% (95% CI 56–80) for highly relevant diagnostic hypotheses ($p < 0.0001$).

4. Discussion

Our findings suggest that a computer simulation fostering free inquiry and a SP simulation lead physicians to similar diagnostic outcomes and clinical decisions. However, while clinicians in the SP simulation do ask essential questions, they also collect less relevant

Table III. Distribution into each simulation method (*n* and percentage of total) of the 1468 distinct items generated in history, physical examination, and decisions phases of the encounters.

		SP		
		Elicited	Not elicited	Total
VIPS	Elicited	416 (28%)	264 (18%)	680
	Not elicited	788 (54%)	na	
Total		1204		1468

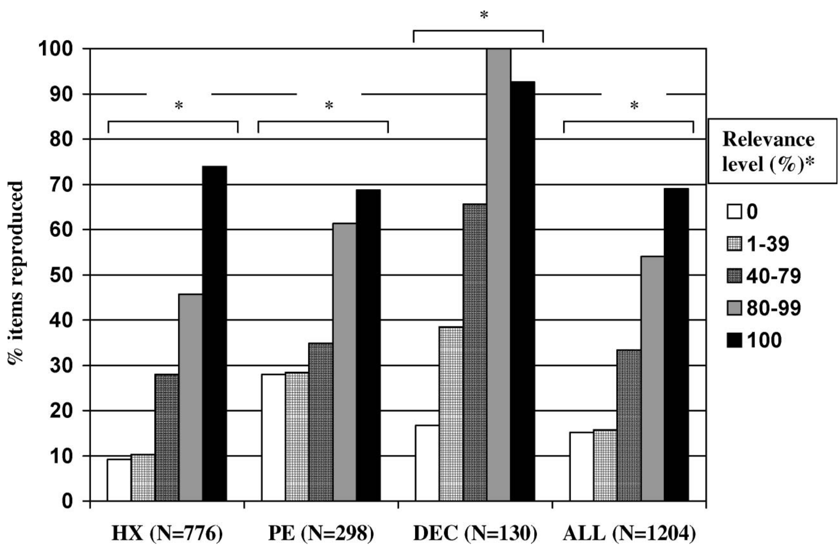

Figure 1. Proportion of items generated in the standardized patient simulation and reproduced in the computer simulation, according to their levels of relevance, for history questions (HX), physical examination procedures (PE), and therapeutic or management decisions (DEC). $*p < 0.0001$ (chi-squared tests). The level of relevance is based on the degree of concurrence (%) among expert internists of a reference group who previously worked up the same cases with a SP: highly relevant items are those elicited by all members (100%) of this reference group.

information to perform a wider exploration of working diagnoses. The computer simulation appears to limit the extent to which physicians collect data and explore working diagnoses. Such a narrower approach may result from the constraints related to the interaction with a machine rather than a person, such as the necessity to type each inquiry. However, this appears to affect more specifically the exploration of secondary working diagnoses than the collection of relevant data, the amount of which remains sufficient to reach the correct diagnoses and make the right decisions. The analysis of the purposes of the history data collected confirms that with the computer simulation, questions aim essentially to characterize complaints and test specific diagnostic hypotheses rather than to screen for the presence of unexpected symptoms.

Numerous reports exist on the use of computer simulation in clinical settings. Among the many publications in this field, the majority are predominantly demonstration or description articles (e.g. 60% of 2763 papers published between 1966 and 1998 [4]). Other studies assess the effectiveness of computer-aided learning to increase specific knowledge and problem-solving skills in various fields (e.g. blood gases [11], emergency [1] or surgery [2,12]). The present study differs from the existing literature by exploring the process of data collection related to the simulation methods. It suggests that a computer simulation allowing for a free inquiry approach accurately makes physicians reproduce their thinking process, as reflected by the information collected and the diagnostic hypotheses explored. This simulation may even foster them to focus early and more specifically on the evaluation of relevant diagnostic hypotheses and the collection of essential data, a condition closely related to diagnostic performance [13]. Research in medical education has shown the importance of an early problem representation [14] and early collection of relevant data framed by relevant diagnostic hypotheses [15,16] in the diagnostic process. This tool, by inviting the subject to systematically consider diagnostic hypotheses, may help to extend the experience and training of learners in clinical settings. Further evaluation should determine whether the use of this kind of simulation will eventually improve their diagnostic performance.

This exploratory study has the strength of analysing the characteristics of the data-collection process and diagnostic hypothesis exploration, but because of the constraints due to this type of assessment and data analysis, the sample size could not be very large, which may limit the interpretation of negative comparisons. Moreover, the generalization of the results to any case and any clinician should be made with caution. Despite this limitation, some characteristics of both simulation methods were sufficiently distinct to be detected. The fact that all the participants solved the cases, first with the SP, then with the VIPS, could raise the issue of a case-recall bias. This is, however, very unlikely, for the following reasons. First, the clinicians received no feedback after the encounters and thus could not memorize the right information to collect. Second, they were not aware that similar complaints could be used in the computer simulation. Third, during the several months that had elapsed between both simulations, they all had regular clinical activities that could interfere with the memorization of the details of each encounter. Finally, care was taken to avoid triggering their memory by changing the context of the patient's history between both simulations, a condition shown to affect the recognition of analogy among problems [8].

5. Conclusion

A clinical work-up performed on a computer simulation allowing for a free-inquiry approach may accurately compare with that observed with a SP simulation in terms of clinical data collection and diagnostic hypothesis exploration. By inciting users to frame their data collection with diagnostic hypotheses, this may promote an early evaluation of relevant working diagnoses and the collection of essential data. Such computerized simulators are not intended to replace SP simulation, particularly with respect to the practice and evaluation of competencies in physical examination, communication, and attitudes, but their contribution to extend the experience and competence of learners in clinical settings should be evaluated further.

Acknowledgements

The authors acknowledge Dr MA Raetzo, MD, Groupe Médical d'Onex, Switzerland, and Dr V Loroch, PhD, Essertines, Switzerland, for their contribution in providing the latest version of the computer simulation program and feeding the software with our cases. The authors also thank the faculty members and residents who so willingly participated in this study. This research was supported by a grant from the University of Geneva (G. and L. Hirsch).

References

1. Weller JM. Simulation in undergraduate medical education: bridging the gap between theory and practice. Medical Education 2004;38:32 – 38.
2. Lattanzi R, Viceconti M, Zannoni C, Quadrani P, Toni A. Hip-Op: an innovative software to plan total hip replacement surgery. Medical Informatics & the Internet in Medicine 2002;27:71 – 83.
3. Ward J, Gordon J, Field M, Lehmann H. Communication and information technology in medical education. Lancet 2001;357:792 – 796.
4. Adler MD, Johnson KB. Quantifying the literature of computer-aided instruction in medical education. Academic Medicine 2000;75:1025 – 1028.
5. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. Academic Medicine 1999;74:S138 – S143.
6. McGee J, Neill J, Goldman L, Casey E. Using multimedia virtual patients to enhance the clinical curriculum for medical students. Medinfo 1998;9:732 – 735.
7. Nendaz MR, Raetzo M, Junod A, Vu N. Teaching diagnostic skills: Clinical vignettes or chief complaints? Advances in Health Sciences Education Theory and Practice 2000;5:3 – 10.
8. Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. Academic Medicine 1998;73:S1 – S5.
9. Charlin B, Desaulniers M, Gagnon R, Blouin D, van der Vleuten C. Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. Teaching and Learning in Medicine 2002;14:150 – 156.
10. Nendaz MR, Gut AM, Perrier A, Reuille O, Louis-Simonet M, Junod AF, Vu NV. Degree of concurrency among experts in data collection and diagnostic hypothesis generation during clinical encounters. Medical Education 2004;38:25 – 31.
11. Vichitvejpaisal P, Sitthikongsak S, Preechakoon B, Kraiprasit K, Parakkamodom S, Manon C, Petcharatana S. Does computer-assisted instruction really help to improve the learning process? Medical Education 2001;35: 983 – 989.
12. Devitt P, Cehic D, Palmer E. Computers in medical education 2. Use of a computer package to supplement the clinical experience in a surgical clerkship: an objective evaluation. The Australian and New Zealand Journal of Surgery 1998;68:428 – 431.
13. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. British Medical Journal 2002;324:729 – 732.
14. Chang RW, Bordage G, Connell KJ. The importance of early problem representation during case presentations. Academic Medicine 1998;73:S109 – S111.
15. Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. Psychological Science 2000;11:112 – 117.
16. Leblanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. Academic Medicine 2002;77:S67 – S69.

Appendix. Excerpts of a transcription of the diarrhoea case (translation)

sequence	question asked	purpose of the question	working diagnosis
history			
1	is it the first time?	characterize the complaint	no hypothesis
2	have you had recent changes in your bowel habits?	test a diagnostic hypothesis	colonic cancer
3	have you abdominal pain?	test a diagnostic hypothesis	colonic cancer
4	at which time in the day?	characterize the complaint	no hypothesis
5	where is the pain?	test a diagnostic hypothesis	chronic pancreatitis
6	how did the pain change since it appeared?	characterize the complaint	no hypothesis
7	what is the character of the pain?	characterize the complaint	no hypothesis
8	what is the frequency of the diarrhoea	characterize the complaint	no hypothesis
9	what is the colour of the stools	test a diagnostic hypothesis	biliary problem
10	did your skin become yellow?	test a diagnostic hypothesis	biliary problem
11	did you see blood in your stools?	test a diagnostic hypothesis	colonic cancer
12	do you suffer from reflux?	test a diagnostic hypothesis	gastric ulcer
13	did your weight change lately?	test a diagnostic hypothesis	colonic cancer
15	can you eat fat meals?	test a diagnostic hypothesis	chronic pancreatitis
16	have you already had abdominal surgery?	test a diagnostic hypothesis	bowel obstruction
17	do you smoke?	routine	no hypothesis
18	what is your medication?	routine	no hypothesis
19	is your familiy in good health?	test a diagnostic hypothesis	cancer
20	did your travel recently?	test a diagnostic hypothesis	chronic pancreatitis
21	have you had fever?	test a diagnostic hypothesis	bowel infection
22	do you have night sweating?	test a diagnostic hypothesis	cancer
23	did you vomit?	test a diagnostic hypothesis	gastric cancer
25	do you drink alcohol?	test a diagnostic hypothesis	liver cancer
26	how much?	characterize the complaint	no hypothesis
physical examination			
27	stethoscope: listen to the abdomen	test a diagnostic hypothesis	intestinal obstruction
28	palpation: epigastric region	test a diagnostic hypothesis	pancreatic cancer/cyst
30	palpation: lower left abdominal quadrant	test a diagnostic hypothesis	colonic cancer
31	palpation: upper right quadrant	test a diagnostic hypothesis	liver cirrhosis
32	palpation: rectal examination	test a diagnostic hypothesis	colonic cancer
33	percussion: abdomen	test a diagnostic hypothesis	ascitis
34	observation: sclera	test a diagnostic hypothesis	biliary problem
35	observation: skin	test a diagnostic hypothesis	biliary problem
36	respiratory frequency	routine	no hypothesis
37	pulsations	routine	no hypothesis
38	blood pressure	routine	no hypothesis
39	temperature	test a diagnostic hypothesis	infectious disease
44	weight	test a diagnostic hypothesis	cancer
decisions			
47	blood tests: coagulation tests	test a diagnostic hypothesis	liver insufficiency
48	blood test: numeration	test a diagnostic hypothesis	digestive blood loss
50	blood test: liver enzymes	test a diagnostic hypothesis	biliary problem
51	blood test: pancreatic enzymes	test a diagnostic hypothesis	acute pancreatitis
52	stools: fat content	test a diagnostic hypothesis	biliary problem
53	stools: culture	test a diagnostic hypothesis	infectious disease
56	new appointment in 5 days		