

## Evaluation of a Decision Support System for Pressure Ulcer Prevention and Management: Preliminary Findings

Rita D. Zielstorff, RN MS\*, Greg Estey, M Ed\*, Amanda Vickery, RN MS+,  
Glenys Hamilton, RN DNSc+, Joan B. Fitzmaurice, RN PhD+, G. Octo Barnett, MD\*

\*Laboratory of Computer Science  
+Department of Nursing  
Massachusetts General Hospital  
Boston, MA 02114

*A decision support system for prevention and management of pressure ulcers was developed based on AHCPR guidelines and other sources. The system was implemented for 21 weeks on a 20-bed clinical care unit. Fifteen nurses on that unit volunteered as subjects of the intervention to see whether use of the system would have a positive effect on their knowledge about pressure ulcers and on their decision-making skills related to this topic. A similar care unit was used as a control. In addition, the system was evaluated by experts for its instructional adequacy, and by end users for their satisfaction with the system. Preliminary results show no effect on knowledge about pressure ulcers and no effect on clinical decision making skills. The system was rated positively for instructional adequacy, and positively for user satisfaction. User interviews related to satisfaction supplemented the quantitative findings. A discussion of the issues of conducting experiments like this in today's clinical environment is included.*

### INTRODUCTION

As part of a research project intended to provide problem-based knowledge to clinicians at the point of care, we developed a system that supports the nurse's development of patient-specific, guideline-based treatment plans for patients who have pressure ulcers or are at risk for developing them. The design and implementation of the system have been described in previous publications<sup>1,2</sup>. The system was implemented experimentally in December 1995 on a 20-bed inpatient orthopedic/neurosurgery unit at the Massachusetts General Hospital. Fifteen of 22 nurses enrolled in the study, and 12 remained enrolled through completion of post-testing. Those who didn't enroll were part-time personnel, or permanent off-shift nurses who felt that the system was mostly intended to help primary nurses with assessment and treatment planning. All of the 15 enrollees on the experimental care unit entered real patient data at least once during the 21-week experimental period.

The control unit was a 28-bed acute orthopedic unit specializing in trauma. Seventeen of 25 permanent RN staff voluntarily enrolled in the study, and 9 of these remained enrolled through completion of post-testing. As an incentive, volunteers on each unit received nominal gift certificates at local stores for each portion of the study protocol they completed.

The evaluation was designed to answer several questions, among them: a) Will the system be acceptable to instructional and content matter experts? b) Will the system improve knowledge about pressure ulcer prevention and treatment? c) Will the system improve clinical decision making skills related to pressure ulcer prevention and treatment? and d) Will the system be acceptable to clinicians who use it? We developed protocols to answer each of these questions. The methods, preliminary results and conclusions for each protocol are presented below.

### PROTOCOL 1. INSTRUCTIONAL ADEQUACY

#### Methods

A one-time survey was completed by three registered nurses who have expertise in the clinical area and/or in instructional technology. The survey instrument used is a slight modification of the Underwood Software Evaluation Tool<sup>3</sup>, a 30-item instrument using Thurstone's equal-appearing interval scaling technique with bipolar descriptors. The items produce scores along four major dimensions: ten items evaluate Content, six items are concerned with Pedagogy, seven items assess Technical Quality and and eight items are concerned with Policy Issues. In addition, each of the experts wrote comments which further explained their ratings.

#### Preliminary Results

The mean scores for each of the four dimensions are presented in Table 1. There was consistently positive scoring among the raters for content and for policy issues, but a marked diversity in scoring for pedagogy and for technical quality of the program. The text comments revealed that the person who gave the low ratings on technical quality of the program had never used a Windows application, and experienced frustration in being expected to know the conventions for interacting with graphical interfaces (such as having to press the tab key to move from field to field in an input form, and having to close a pop-up window before trying to interact with its parent). The same person produced the lowest mean score on the pedagogy dimension.

Three of the thirty individual items on the survey received a unanimous score of +3. These were: "well balanced and representative information is

presented”; “the software package is compatible with the goals of the clinical training program”; and “the software package fits in well with other instructional materials already being used in clinical areas”. The lowest-scoring individual item, which had a mean of -1 among the three raters, was for the item which described the ability of the program to be used by two or more users interacting with each other (the application was not designed to do this).

The reviewers appended a total of 31 comments to their survey forms. Sixteen comments were specific suggestions for the clinical content or the decision rules. Nine comments were criticisms of the system behavior or interface design, or questions regarding the intent of some part of the system. Six comments praised the system or some specific aspect of it.

### **Conclusion**

Although there was some diversity in the ratings of the three evaluators, the program received an overall positive evaluation as reflected in both the survey form items and in the text comments. Lack of familiarity with graphical interfaces influenced the ratings of one of the evaluators with regard to technical quality of the program.

## **PROTOCOL 2. IMPACT ON KNOWLEDGE**

### **Methods**

Knowledge was measured with a 30-item, computer-based multiple choice test developed for this project following classic methods of test construction. An initial set of 92 items was culled through expert review and pretesting to thirty. The items ranged from easy to difficult, with adequate discrimination power and consideration of the representation of all content areas. The test was administered to the volunteer subjects from the experimental unit and the control unit. Both groups completed the test prior to implementation of the system on the experimental unit; both groups completed the test again (with the same questions but asked in a different order) after the system had been implemented on the experimental unit for 21 weeks. Because of attrition, pre- and post- paired scores are available for 12 of the 15 original enrollees on the experimental unit, and 9 of the 17 original enrollees on the control unit.

### **Preliminary Results**

Table 2 summarizes the results. There was no appreciable difference between the scores of nurses who used the system on the experimental unit vs. those who did not on the control unit; using a paired-samples comparison, nurses on the experimental unit showed no appreciable difference in knowledge scores after using the system.

### **Conclusion**

A 21-week exposure to the system had no effect on nurses' knowledge of pressure ulcer prevention and treatment, as reflected in the knowledge test.

## **PROTOCOL 3. IMPACT ON CLINICAL DECISION MAKING**

### **Methods**

To test clinical decision making a computer based simulation program pertaining to pressure ulcer prevention and treatment was developed. Case simulations describing a patient scenario were constructed and reviewed by expert clinicians in skin care management. Scoring is based on: a) Information collection (completeness and efficiency); b) Identification of presence of pressure ulcer; c) Selection of pressure ulcer etiology(ies); d) Identification of risk factor presence; e) Selection of risk factor etiology (ies); and f) Selection of treatments.

The program has three cases, a sample case and two test cases. The sample case gives the participant an opportunity to become familiar with the format and to receive a case analysis and a comprehensive analysis of performance. This analysis is not given after the two test cases. The sample case may be repeated. The sample case and the two test cases take approximately 45 minutes to complete.

All enrolled subjects from the experimental and control units completed the three computer-based simulations prior to implementation of the program on the experimental unit. The cases were completed again by 13 of the 15 original enrollees on the experimental unit (87%), and 9 of the original 17 enrollees on the control unit (53%) after the 21-week experimental period.

### **Preliminary Results**

Table 3 summarizes the results of the preliminary analysis. These results are similar to those on the knowledge test. The control group's performance declined or stayed the same by most measures; the experimental group's performance varied slightly in both directions, but generally stayed the same.

### **Conclusion**

A 21-week exposure to the system had no effect on nurses' clinical decision making related to pressure ulcer prevention and treatment, as reflected in the case simulations. The small number of participants, and the attrition of 47% of the control group, make definitive conclusions impossible.

## **PROTOCOL 4. USER SATISFACTION**

### **Methods**

End-user satisfaction was assessed at the end of the experimental period from the 15 volunteer subjects, all of whom worked on the experimental unit and used the system at least once. Both quantitative and qualitative approaches were used. First, the clinicians were given a survey form to complete. A

**Table 1. Ratings of Instructional Adequacy\***

|                   | Rater 1 | Rater 2 | Rater 3 | Mean  |
|-------------------|---------|---------|---------|-------|
| Content           | +2.25   | +2.40   | +2.60   | +2.41 |
| Pedagogy          | +2.67   | +0.33   | +3.00   | +2.00 |
| Technical quality | +2.83   | -0.83   | +2.33   | +1.44 |
| Dimension         | +2.00   | +1.00   | +1.00   | +1.66 |

\* Possible range: - 3 = lowest, + 3 = highest

**Table 2. Knowledge Scores\***

|        | Experimental |      | Control |      |
|--------|--------------|------|---------|------|
|        | Pre          | Post | Pre     | Post |
| Mean   | 18           | 19   | 20      | 19   |
| Median | 18           | 18   | 21      | 19   |
| Low    | 15           | 15   | 14      | 13   |
| High   | 21           | 23   | 23      | 25   |

\* Possible range: 0-30

**Table 3. Clinical Simulation Scores**

|  | Experimental (n=13) |      | Control (n=9) |      |
|--|---------------------|------|---------------|------|
|  | Pre                 | Post | Pre           | Post |
| <b>Case 1</b>                          |                     |      |               |      |
| Diagnosis correct                      | 7                   | 8    | 4             | 5    |
| Proportion correct etiology (mean)     | .64                 | .58  | .59           | .48  |
| Risk assessment correct                | 11                  | 12   | 9             | 9    |
| Proportion correct risk factors (mean) | .78                 | .89  | .70           | .67  |
| Proportion correct therapies (mean)    | .44                 | .44  | .54           | .47  |
| <b>Case 2</b>                          |                     |      |               |      |
| Diagnosis correct                      | 12                  | 11   | 9             | 9    |
| Proportion correct etiology (mean)     | N/A                 | N/A  | N/A           | N/A  |
| Risk assessment correct                | 7                   | 8    | 4             | 4    |
| Proportion correct risk factors (mean) | .58                 | .42  | .44           | .44  |
| Proportion correct therapies (mean)    | .58                 | .56  | .63           | .56  |

**Table 4. End-User Satisfaction: Mean ratings for individual items\***

| Item   | Mean | (s.d.) |
|--|------|--------|
| 1. Does the system provide the precise information you need?               | 3.93 | (0.96) |
| 2. Does the information content meet your needs?                           | 3.66 | (1.04) |
| 3. Does the system provide displays that seem to be exactly what you need? | 3.53 | (1.18) |
| 4. Does the system provide sufficient information?                         | 4.20 | (0.67) |
| 5. Is the system accurate?   | 4.13 | (0.74) |
| 6. Are you satisfied with the accuracy of the system?                      | 4.26 | (0.59) |
| 7. Do you think the output is presented in a useful format?                | 4.33 | (0.89) |
| 8. Is the information clear?   | 4.33 | (0.81) |
| 9. Is the system user friendly?  | 4.28 | (0.91) |
| 10. Is the system easy to use?   | 3.93 | (1.16) |
| 11. Do you get the information you need in time?                           | 3.93 | (0.79) |
| 12. Does the system provide up-to-date information?                        | 4.26 | (0.59) |

\*Scoring: 1- Almost never, 2- Some of the time, 3- About half of the time, 4- Most of the time, 5- Almost always

**Table 5. Mean scores of Components of End User Satisfaction (1=lowest, 5=highest)**

| Component   | Mean Score |
|-------------|------------|
| Content     | 3.83       |
| Accuracy    | 4.20       |
| Format      | 4.33       |
| Ease of Use | 4.06       |
| Timeliness  | 4.10       |

12-item questionnaire developed by Doll and Torkzadeh<sup>4</sup> was distributed. The instrument utilizes a five-point Likert-type scale to quantify users' perceptions of five system components: Content, Accuracy, Format, Ease of Use, and Timeliness. There are four questions assessing the dimension of Content; two questions assessing Accuracy; two questions assessing Format; two questions assessing Ease of Use, and two questions assessing Timeliness.

We selected this instrument because of its documented reliability and validity, because it assessed parameters that were appropriate for our application, and because it was brief enough to be practical in our environment.

Of several tools we reviewed, this was the only one that met all of these criteria. In addition to the written survey, face-to-face interviews were conducted with the same group of clinicians.

The developers of the End-User Computing Satisfaction questionnaire used the form to evaluate a range of applications in 44 firms representing various industries. The industries included health care, the range of applications included decision support systems, and the categories of workers included Professional Employees Without Supervisory Responsibilities (typical of our users). Over 600 users responded to those surveys. Out of this experience, the following statistics were reported: the mean score among applications was 49.09; median, 51; minimum, 16; maximum, 60; standard deviation, 8.302. In addition, percentile scores were reported: A score of 48 was in the 40th percentile, and 51 was in the 50th percentile.<sup>5</sup> We will refer to these statistics in reviewing the results of our own survey.

#### **Preliminary Results**

All fifteen users of the experimental system returned the survey. The mean ratings for each of the twelve questions in the survey are itemized in Table 4. The mean score for each of the dimensions is shown in Table 5. The mean total score for the survey was 48.67 out of a possible 60, with a standard deviation of 6.27. This is similar to the population statistics reported by Doll & Torkzadeh. The average total score of 48.6 for the Pressure Ulcer Prevention and Management System falls in approximately the 40th percentile of total scores for all applications surveyed in the Doll & Torkzadeh study.<sup>5</sup>

The face-to-face interviews were consistent with the survey results. The lowest-scoring dimension was Content. There were several comments pertaining to content, such as "I would rather have put in basic information about my patient and then have it give me a simple recommendation," and "It's too specific to pressure ulcers...it would be better if it had general wound care also," and "I didn't feel I got the treatment knowledge that I needed." The highest-scoring dimension was Format, which is reflected in such comments as "Easy to read," "Nice to have a

printed form [treatment plan]," and "Very orderly and logical." There were no comments on accuracy of the content. Related to Ease of Use, there were comments such as "Just as easy as writing," and "It was easy when I did it."

#### **Conclusion**

The system received an overall positive rating by end users, both on the written structured survey and in face-to-face interviews.

## **DISCUSSION**

A set of protocols for evaluating the impact of the Pressure Ulcer Prevention and Management System has shown mixed results thus far. Written ratings, textual comments, and structured interviews with users and with content experts yielded generally positive results, but there is no evidence that the system has influenced more distant outcomes such as knowledge and clinical decision making. There are many possible reasons for this, some having to do with the experiment itself, and some having to do with the environment in which it was conducted.

It is possible, for example, that the length of time that the subjects were exposed to the intervention was not long enough to have measurable effects. Or it may be that some of the outcomes are too distant from the intervention to have been influenced, or that the instruments were not sensitive enough to measure the changes. We know from direct observation that usage among the subjects was uneven and in some cases infrequent, due to unexpectedly low census during the experimental period. We also know that nurses on both units were subjected to many environmental stresses during the experimental period, with beds closing, and staff often in danger of being transferred or even laid off. Obtaining their cooperation and stimulating their continued commitment was extremely difficult under these conditions.

With regard to the experiment itself, the small number of participants makes it difficult to make clear judgments about the results of the quantitative measures. The high attrition rate of subjects on the control unit, despite persistent efforts on the part of our research assistants, further limits our understanding of the results.

We were able to draw some conclusions from our observations of the nurses' behavior on the experimental unit, and from the comments they made during the interviews. A common theme, for instance, was that there was not enough gain for the effort required to enter the data into the system. In today's world, clinicians need *rapid* access to *very specific* knowledge related to a particular problem at hand. In our system design, we required the nurses to use a comprehensive assessment module, provided

advice on diagnosis only when it was corrective, and encouraged a comprehensive plan of care.

We are redesigning the system to convert the procedural rules imbedded in the system to a set of problem-specific algorithms that we will make available over the World Wide Web (see related poster by Hulse et al.<sup>6</sup>). In this way, clinicians can seek answers to their specific questions on an as-needed basis, with data entry focused only on information needed to supply an answer to the specified problem. Knowledge will thus be provided “just-in-time” to influence the decision at hand.

#### Acknowledgments

This work was supported by Grant 5 R18 HS06575, Agency for Health Care Policy and Research; by Grant 5 R01 LM05200, and Grant 1 T15 LM07092, National Library of Medicine; and by an educational grant from Hewlett Packard Corporation.

#### References

1. Zielstorff RD, Barnett GO, Fitzmaurice JB, et al. A Decision Support System for Prevention and Treatment of Pressure Ulcers Based on AHCPR Guidelines. In Cimino, J (ed.). Proceedings of the 1996 AMIA Annual Fall Symposium. Phila: Hanley & Belfus, 1996, pp. 562-566.
2. Estey G, Shahzad C, Zielstorff R, et al. A Demonstration of Integrated Access to Pressure Ulcer Guidelines. In Cimino J (ed). Proceedings of the 1996 AMIA Annual Fall Symposium. Phila: Hanley & Belfus, 1996, p. 922.
3. Underwood SM. Measuring the validity of computer-assisted instructional media. In: OL Strickland & CF Waltz (eds). Measurement of Nursing Outcomes. Vol 2. NY: Springer Publishing, 1988:295-313.
4. Doll WJ, & Torkzadeh G. The Measurement of End-User Computing Satisfaction. MIS Quart 1988, 12:259-274.
5. Ibid, p. 270.
6. Hulse M, Zielstorff RD, Estey G, et al. Design Considerations for User Interface to an Expert System on the World Wide Web. (these proceedings).