

Accuracy of a Web-based System for Monitoring Chronic Wounds

HYUNGJIN MYRA KIM, Sc.D., JULIE C. LOWERY, Ph.D., JENNIFER B. HAMILL, M.P.H.,
and EDWIN G. WILKINS, M.D., M.S.

ABSTRACT

This study evaluated the accuracy of a store-and-forward telemedicine system for assessing the status of chronic wounds, including those surgically repaired. Digital photos and other patient and wound data were collected by a nurse using a laptop and transmitted via the Internet to a database, which organized and posted the data onto a web page for access by the telemedicine physician. Two Veterans' Affairs (VA) medical centers and two specialties (plastic surgery, physical medicine and rehabilitation) participated in the study. Study patients included inpatients and outpatients with pressure ulcers of stage II, III, or IV, plus outpatients with diabetic foot ulcers or venous stasis ulcers. All patients were assessed both in-person (the "gold standard") and with the telemedicine system using yes/no responses and a 5-point scale, respectively, on four diagnostic questions concerning wound healing and infection, based on AHCPR guidelines. A total of 70 patients were enrolled, with data collected on 430 visits: up to 6 visits per wound. Percentage agreement for all visits ranged from 67.1 for "not healing" to 88.8 for "cellulitis present." Sensitivity ranged from 0.32 for cellulitis to 0.63 for necrosis; and specificity ranged from 0.80 for necrosis to 0.91 for cellulitis. Although agreement of the telemedicine system was not high, it was not significantly less than interphysician agreement on in-person assessments. A relatively inexpensive store-and-forward telemedicine system for monitoring the status of chronic wounds has the potential to improve access to specialty care for patients who are not currently receiving routine monitoring by specialized nurses or physicians.

INTRODUCTION

PRESSURE ULCERS are a significant and costly problem in elderly and disabled patients.^{1,2} One of the contributors to the cost is their lack of access to timely, specialized care. The use of a telemedicine system could improve access to providers with specialized training, by enabling providers to assess the status and man-

age the treatment of ulcers without personally seeing the patients. Increased access to specialized care should reduce the incidence and improve the treatment of infected ulcers, which should, in turn, reduce the costs of treating patients with these conditions. The costs of care include the outpatient visits and hospitalizations that could be prevented with the provision of more timely care, as well as the costs of

VA Center for Practice Management and Outcomes Research and Section of Plastic and Reconstructive Surgery, University of Michigan, Ann Arbor, Michigan.

transporting disabled patients to a hospital or physician's office. Hospitalization days of patients who undergo surgery for repair of their wounds could be reduced with telemedicine by enabling patients to be monitored postoperatively at home or in an extended care facility.

The primary objective of this study was to evaluate the clinical accuracy of a store-and-forward telemedicine system for assessing the status of different types of ulcers, both chronic and those surgically repaired. The purpose of most telemedicine systems, including the one evaluated here, is to enable clinicians to provide a diagnosis for patients at a remote site as an alternative to seeing the patient in-person. To evaluate the clinical accuracy of the proposed telemedicine approach, therefore, the results of the assessment process using telemedicine are compared with results under the current standard of care based on an in-person assessment.

A number of systems for providing wound care at a distance have been described in the medical literature. Several Veterans' Affairs (VA) medical centers use videoconferencing to manage pressure ulcers and other conditions of patients with spinal cord injuries.^{3,4} However, the equipment for these systems can be expensive and occasionally unreliable. The Shepherd Center in Atlanta uses a less expensive, telephone-based system to monitor the wounds of spinal-cord-injured patients in their homes.^{5,6} The Picasso Still-Image Videophone allows the real-time transmission of still images of diagnostic quality from patient to clinician over the telephone while they converse.

Among the lessons learned from telemedicine demonstrations in the military thus far is that lower-cost store-and-forward data transmission, rather than expensive real-time video, is adequate for most transactions.⁷ Because of their relatively low cost and ease of use, the majority of telemedicine applications in the future are predicted to be store-and-forward and Internet-based. Store-and-forward applications involve digital photographs and other clinical data being captured and transmitted to the telemedicine physician. The physician can then view the data via e-mail, a web site, or a file on a shared server at her or his convenience. Only a few examples of store-and-forward systems

for conducting evaluations of wounds are described in the literature. The British Defense Medical Services have used digital cameras and electronic mail to transmit wound images between a hospital in Bosnia and a British Defense hospital in the United Kingdom.⁸ Wirthlin et al. described a store-and-forward system for the remote management of vascular wounds.⁹

MATERIALS AND METHODS

Study design

The study was a prospective cohort design comparing the diagnostic evaluation of a wound by a treating physician (in-person assessment) with a diagnostic evaluation by a remote physician using the telemedicine system. The in-person physicians' diagnoses are considered the "gold standard" or "true diagnosis" for calculating accuracy, which is the standard approach taken in similar studies evaluating the accuracy of telemedicine diagnoses, such as studies of teledermatology.¹⁰⁻¹³ Two VA medical centers (Ann Arbor, MI, and Augusta, GA) participated in data collection. Participants were hospital inpatients, outpatients, or nursing home residents with the following types of wounds: chronic stage II, III, or IV pressure ulcers; patients who had undergone a skin flap procedure for treatment of a grade III or IV pressure ulcer; diabetic foot ulcers; or venous stasis ulcers. Patients with multiple wounds were included, and mentally incompetent patients were excluded.

Six participating physicians worked in pairs and rotated roles within each pair as the in-person and telemedicine evaluator as their schedules allowed, so that each physician served an approximately equal amount of time in each role. One pair of plastic surgeons participated at the Ann Arbor site reviewing data for all chronic wounds. In Augusta, a pair of physical medicine and rehabilitation (PM&R) specialists from the Spinal Cord Injury (SCI) Unit reviewed chronic wounds, while plastic surgery residents from the Medical College of Georgia reviewed postoperative wounds. The plastic surgery residents were in their final year of postgraduate plastic surgery training and had

considerable experience caring for patients with both acute and chronic wounds. If a patient with an open wound underwent a surgical closure while participating in the study in Augusta, the pair of physicians reviewing the patient's data would switch from PM&R to plastic surgery. Each wound was assessed by an in-person and telemedicine physician, and the physicians were blinded to each other's assessments. Data collection began in March, 1999, and concluded in August, 2000.

Data collection

Potential patients were identified by a nurse coordinator, who was funded by the study at each site. The Augusta VA Medical Center (VAMC) has a regional SCI center and the nurse coordinator received notification of all new or recurring ulcers on a regular basis from the participating physicians. At the Ann Arbor VAMC, patients were recruited from several locations within the medical center. The nurse coordinator had access to the computer-based scheduling system and also received notification from cooperating nurses and physicians. At both sites, after identifying a potential subject, the nurse coordinators approached a patient about participation and verbally explained the study. If the patient was interested, he was given study materials, including an informed consent document and study questionnaire, and participating patients were reimbursed at the rate of \$10 per data collection session.

The nurse coordinators collected patient and wound data at each study visit for the telemedicine evaluation. Data on each study patient were collected for a maximum of six visits, or until the wound was healed or the patient was discharged, whichever came first. The telemedicine system for wound assessment consisted of three major data collection components: digital photographs of the ulcer; quantitative measurements of the ulcer (i.e., ulcer area and volume); and other ulcer and patient data collected by the nurse (e.g., wound drainage, patient mobility, continence status, nutrition status, etc.). The specific data collection components were based on the Agency for Health Care Policy and Research (now the Agency for

Health Research and Quality) guidelines on treatment of pressure ulcers.¹⁴ Information on the system's data collection protocol and technical specifications are published elsewhere¹⁵ or can be viewed at http://www.hsrp.ann-arbor.med.va.gov/telemed_toc.htm.

The nurse entered the data into an Access database, from which they were uploaded periodically via an Internet connection to the study's central Oracle database on a remote server at the University of Michigan. The digital images were uploaded to an image directory. Web software (ColdFusion) downloaded the images from the directory and the data from the Oracle database into tables, graphs, and images on the web site. The password-protected web site was accessible to the telemedicine physician from any location where Internet access was available. Patient confidentiality was maintained by use of study identification numbers for each patient: no patient names, medical record, or Social Security numbers were presented. This precaution, plus the use of a password to access the site, precluded the need for encryption software. A demonstration version of the Web page can be viewed at www.wats.ann-arbor.med.va.gov/ (user ID = Demo, password = DMWound).

In-person and telemedicine diagnostic variables

The in-person physician personally examined the ulcer, reviewed the patient's medical record, recorded his or her observations and diagnoses in the medical record, and responded to four or five questions on a study form. Consistent with the AHCPR guidelines¹⁴, the diagnostic questions for chronic ulcer patients were as follows: Is necrotic tissue present? Is cellulitis suspected? Is the wound getting smaller? Is osteomyelitis suspected? A fifth question was added for patients with pressure ulcers who had undergone operative repair: Is the wound closed? For all diagnostic questions, the in-person physicians were required to answer the questions either yes or no, since their answers provided the standard against which the telemedicine assessments were compared. The telemedicine physicians reviewed the ulcer and patient data on the web site and then responded to the same assessment questions us-

ing a five-point scale of "positively yes," "probably yes," "not sure," "probably no," and "positively no." For each visit, the telemedicine physicians entered their responses into a page on the web site.

Training

To establish the level of agreement between physicians performing conventional (i.e., in-person) assessments, and to compare this to the rate of agreement observed between the in-person and telemedicine assessments obtained from the study, a baseline rate of agreement between the participating physician pairs was established at both sites. Each physician pair saw a group of pressure ulcer patients, different from the study participants, in person and independently answered the five diagnostic questions. In addition, the study physicians participated in an in-service teleconference on the telemedicine review and assessment process with the investigators. Each step of the web process was reviewed and discussed using data from a sample of wounds before the study began.

The two nurses participated in a hands-on training session at the Ann Arbor VAMC, conducted by a consultant with experience in the data collection protocol. Both nurses took digital photographs and measurements of ulcer volume (using a potassium alginate compound to take a mold of the wound) and area (using a computer program to trace the edge of the wound in the digital photograph) on the same patients. Results for both nurses for each patient were compared, and the trainer and nurses worked together to refine the protocol in cases where differences existed.

Because the digital photographs provide most of the information used by the telemedicine physicians in making their assessments, particular attention was paid to ensuring the quality of the photos, by standardizing lighting and orientation. Differences in orientation were assessed as part of the evaluation of variability in ulcer area, because area changes with the orientation of the photo. Assessment of differences in lighting could not be performed quantitatively; therefore, the trainer and nurses used professional judgment to review the pho-

tographs for differences in lighting and coloration and to identify means of reducing variability. In addition to the training provided prior to implementation, refinements to the digital photo protocol continued on an as-needed basis. A senior medical photographer from the Ann Arbor site consulted on the digital camera's operation at several times in the study when room lighting and patient positioning posed problems.

Data analysis

Data were collected on up to six visits and five wounds per patient, and these data were used for two sets of analyses: first visits only and all visits. The unit of analysis was the wound, and multiple wounds of any given patient were considered independent. When all visit data were included, statistical adjustment was made whenever possible for the potential correlation within multiple assessments made for any one wound. Agreement between the in-person and telemedicine assessments was measured using percentage agreement and kappa. Percentage agreement is one of the most commonly reported measures in other studies assessing the accuracy of telemedicine assessments. The kappa statistic was also used because it adjusts for agreement due to chance.¹⁶ Values of kappa can range from -1 to 1 , depending on the strength of agreement; positive values indicate agreement, the value of 0 indicates no agreement, and negative values indicate disagreement. To assess agreement, the five-point telemedicine assessments were compressed from "positively yes" or "probably yes" into "yes," and from "positively no," "probably no," or "not sure" into "no."

The wound evaluations included in this study consisted of three combinations of physician and wound types: chronic wounds evaluated by PM&R physicians (chronic/PM&R), chronic wounds evaluated by plastic surgeons (chronic/plastic), and repaired wounds evaluated by plastic surgeons (repaired/plastic). To assess differences in percentage agreement among the three subgroups, chi-square test was used for the analysis of first visit data. For all visit data, because of the potential correlation

within patients, differences in agreement rate among the three subgroups were evaluated by modeling agreement as a binary response (agreed = 1; not agreed = 0) and using a random-effects logistic regression model with patient as a random effect and with two indicators for the three subgroups of physician/wound type.¹⁷ The random-effects model accounts for potential correlation within patient. To test for differences in agreement rate across the physician/wound type subgroups, a likelihood ratio test was constructed by comparing a model with the group indicators and a model without the group indicators.

As measures of accuracy, sensitivity (proportion of in-person positive assessments correctly identified by the telemedicine assessments) and specificity (proportion of in-person negative assessments correctly identified by the telemedicine assessments) were calculated. Because we did not obtain definitive diagnoses (i.e., through laboratory confirmation) of cellulitis and osteomyelitis, we recognize that we are not truly assessing accuracy of the telemedicine assessments. However, our objective was to compare the telemedicine assessments to the in-person assessments as if the in-person assessments were definitive, because the in-person assessments are the standard of care against which we wish to evaluate the telemedicine system. As measures of accuracy, sensitivity and specificity are useful for understanding the nature and potential implications of disagreement (i.e., to determine if disagreement is due more to the failure to identify problems or to the failure to identify healing).

For first visit data, sensitivity and specificity were calculated as proportions and physician/wound subgroup comparisons were made using a chi-square test. For all visit data, sensitivity and specificity were calculated using an ordinal regression model, which is a flexible way to model the accuracy data and gives robust standard errors by allowing for potential within subject correlation.¹⁸ The physician/wound subgroup comparisons, however, were made using chi-square statistics assuming independence, which were suggested by our data and seemed reasonable; the within-person correlation is likely to be not substantial, because the gold standard status of

any given wound can also change over the visits.

As a summary measure of diagnostic accuracy, the area under the receiver operating-characteristic curve (AUROC) was also calculated. AUROC combines both the sensitivity and specificity while varying the cutoff for classifying positive versus negative diagnoses.¹⁹ An AUROC of 1 indicates perfect diagnostic accuracy, while an AUROC of 0.5 indicates no diagnostic power. For first-visit data, the ROC curves were estimated nonparametrically, and the area was calculated using a trapezoidal rule.²⁰ The 95% confidence intervals for AUROC were calculated, and the AUROC among physician/wound type subgroups were compared using a nonparametric method. When all visit data were included, statistical adjustment was made for the potential correlation within patients by utilizing an ordinal regression model to obtain AUROC.²¹ The physician/wound type subgroup differences in the AUROC were tested assuming independence of multiple assessments of wounds of a person, however.

RESULTS

Sample description

Seventy patients were recruited to the study, and of those, one did not complete a baseline study questionnaire. All 69 patients who completed a questionnaire were male with a mean (range) age of 59 (range, 24–83) years, and 35.3% (24/68) were married or had a live-in partner. In considering the potential applicability of the telemedicine system to home care, the majority of patients (97.1%, 67/69) lived at home rather than in a nursing home. Their living situation varied, with 41.3% (26/63) living without assistance, while the majority (58.7%) received some kind of assistance or care at home (39.7% a full- or part-time caregiver, 12.7% some assistance, and 6.3% a full-time nurse). In addition, the majority (63.3%) of participating patients considered their overall health to be good or very good. There were no significant differences between the two participating sites in the demographics of the sample.

TABLE 1. DISTRIBUTION OF WOUNDS BY TYPE AND NUMBER (PERCENT)

	Wound type					
	Pressure ulcers			Post-op closures	Vascular ulcers	Total
	Stage 2	Stage 3	Stage 4			
No. (%) of wounds	9 (7.6)	16 (13.6)	44 (37.3)	19 ^a (16.1)	30 (25.4)	118 (100.0)
Ann Arbor	7 (14.0)	5 (10.0)	10 (20.0)	0 (0.0)	28 (56.0)	50 (100.0)
Augusta	2 (2.9)	11 (16.2)	34 (50.0)	19 ^a (27.9)	2 (2.9)	68 (100.0)

^aIncludes 10 wounds that were originally chronic pressure ulcers and then were surgically repaired.

One hundred and eight wounds were included in the study; but for the 10 chronic wounds that were later surgically repaired, the postsurgical visit data were treated as those of a new wound. For the statistical analyses, therefore, the study was considered to have data on 118 wounds. Of the 118 wounds, 99 were chronic ulcers and 19 were surgically repaired ulcers. Table 1 presents the distribution of wound types, which varied significantly ($p < 0.0001$, chi-square analysis) across the two study sites.

A total of 430 wound assessments were made, and of those, 27% (118/430) were first

visits, decreasing steadily to 9% (39/430) that were the sixth (and final) visit. The average time between visits was 9.3 days (standard deviation = 8.5, min = 2, max = 56).

Prevalence

Table 2 shows the prevalence (percentage of positive assessments made by in-person physicians) of each of the wound condition assessments, for all visits and for first visits. The prevalence of cellulitis was low for all ulcers and for the individual subgroups. For repaired ulcers, the prevalence was low for all of the conditions. When prevalence is very low, sensitiv-

TABLE 2. PERCENT (n/N) OF WOUNDS WITH THE CONDITION AS ASSESSED BY THE IN-PERSON PHYSICIAN (TRUE POSITIVES)

Ulcer/physician	Condition	All visits	First visit
All ulcers	Not healing	32.6 (97/298)	45.9 (39/85) ^a
	Necrosis	25.4 (109/430)	38.1 (45/118)
	Cellulitis	4.2 (18/430)	5.1 (6/118)
Repaired ulcer	Not healing	10.0 (7/70)	5.9 (1.17) ^a
	Necrosis	7.6 (7/92)	0 (0/19)
	Cellulitis	4.4 (4/92)	0 (0/19)
	Not closed	4.4 (4/92)	0 (0/19)
Chronic ulcer	Not healing	39.5 (90/228)	55.9 (38/68) ^a
	Necrosis	30.2 (102/338)	45.5 (45/99)
	Cellulitis	4.1 (14/338)	6.1 (6/98)
	Osteomyelitis	45.7 (138/302)	49.4 (44/89)
Chronic ulcer by PM&R	Not healing	37.1 (49/132)	56.41 (22/39) ^a
	Necrosis	21.0 (39/186)	30.6 (15/49)
	Cellulitis	0.5 (1/186)	2.0 (1/49)
	Osteomyelitis	61.5 (99/161)	72.1 (31/43)
Chronic ulcer by plastic surgeon	Not healing	42.7 (41/96)	55.2 (16/29) ^a
	Necrosis	41.5 (63/152)	60.0 (30/50)
	Cellulitis	8.6 (13/152)	10.0 (5/50)
	Osteomyelitis	27.7 (39/141)	28.3 (13/46)

^aThe assessment "Is the wound getting smaller?" requires that the physician has seen the patient on at least one previous occasion to make a comparison. Therefore, these assessments were from the second visit.

ity and, thus, AUROC, cannot be estimated reliably or cannot be estimated at all. Therefore, subgroup analyses by wound type and physician group were limited by the lack of sample size for those truly having the conditions.

Agreement

Table 3 shows percentage agreement and kappa for measuring agreement between the telemedicine and in-person assessments of all visit data and first visit data. Percentage agreement on assessments for first visits ranged from 60.0 for not healing to 100.0 for wound not closed (repaired ulcers). Percentage agreements for all visit data tended to be similar or higher than those of first visit data, with the exception of osteomyelitis, in which agreements for all visit data were slightly lower. Except for osteomyelitis, significant ($p < 0.05$) or margin-

ally significant ($p < 0.10$) differences in percentage agreement among the three wound/physician subgroups were observed, in which agreement tended to be greater for assessments made by the plastic surgeons who reviewed the repaired wounds than for the physicians who reviewed the chronic wounds. The repaired wounds, however, had very low prevalence of true problems (refer to Table 2), so their agreement rates were based on considerably more wounds without problems than wounds both with and without problems. High agreement likewise tended to occur when the prevalence of cellulitis was low, especially cellulitis for chronic ulcers evaluated by PM&R physicians.

Although the kappa statistics for all assessments combined showed agreement better than that by chance, they were generally low. Kappa values for all visit data were similar to those of

TABLE 3. PERCENT AGREEMENT AND KAPPA STATISTICS MEASURING AGREEMENT BETWEEN TELEMEDICINE AND IN-PERSON ASSESSMENTS OF ALL VISIT AND FIRST VISIT^a DATA^b

		% Agreement (n/N)		Kappa (p-value ^c)	
		All visits	1st visit only	All visits ^d	1st visit only
Not healing	All	67.1 (200/298)	60.0 (51/85)	0.22 (<0.01)	0.19 (0.04)
	Chronic/PM&R	61.4 (81/132)	51.3 (20/39)	0.12 (0.08)	0.05 (0.36)
	Chronic/plastic	57.3 (55/96)	58.6 (17/29)	0.14 (0.09)	0.16 (0.19)
	Repaired/plastic	91.4 (64/70)	82.4 (14/17)	0.45 (<0.01)	-0.09 (0.65)
	<i>p</i> -value ^e	<0.01	0.09	0.89	0.65
Necrosis	All	77.0 (331/430)	74.6 (88/118)	0.43 (<0.01)	0.47 (<0.01)
	Chronic/PM&R	69.9 (130/186)	65.3 (32/49)	0.17 (<0.01)	0.17 (0.12)
	Chronic/plastic	75.7 (115/152)	76.0 (38/50)	0.51 (<0.01)	0.49 (<0.01)
	Repaired/plastic	93.5 (86/92)	94.7 (18/19)	0.47 (<0.01)	g
	<i>p</i> -value ^e	<0.01	0.04	<0.01	0.11
Cellulitis	All	88.8 (382/430)	81.4 (96/118)	0.18 (<0.01)	0.15 (0.02)
	Chronic/PM&R	95.7 (178/186)	91.8 (45/49)	-0.01 (0.58)	-0.03 (0.60)
	Chronic/plastic	79.0 (120/152)	66.0 (33/50)	0.21 (<0.01)	0.12 (0.12)
	Repaired/plastic	91.3 (84/92)	94.7 (18/19)	-0.05 (0.67)	g
	<i>p</i> -value ^e	<0.01	<0.01	f	f
Osteomyelitis	All chronic	72.5 (219/302)	75.3 (67/89)	0.44 (<0.01)	0.50 (<0.01)
	Chronic/PM&R	68.9 (111/161)	72.1 (31/43)	0.37 (<0.01)	0.42 (<0.01)
	Chronic/plastic	76.6 (108/141)	78.3 (36/46)	0.32 (<0.01)	0.38 (<0.01)
	<i>p</i> -value ^e	0.19	0.50	0.65	0.81
Not closed?	Repaired/plastic	96.7 (89/92)	100.0 (19/19)	0.55 (<0.01)	g

^aIncludes the first visit after a surgical repair for initially chronic patients who were later surgically repaired. For "not healing," analyses were conducted on the second visit.

^bTelemedicine assessments of "positively yes" or "probably yes" were categorized as "yes"; telemedicine assessments of "positively no," "probably no," or "not sure" were categorized as "no."

^cTest for kappa > 0.

^dFor all visit data, kappas were calculated ignoring any potential within-patient correlation.

^eFor agreement rates, *p*-values test for no difference in agreement across the three physician and wound type combinations. For kappa, *p*-values test for no difference in kappa values between the two chronic ulcer subgroups only, because low prevalence of conditions in the repaired wound subgroup makes the kappa values unreliable.

^f*P*-value was not calculated because of low prevalence in at least one subgroup.

^gKappa was not calculated because no wounds had this condition.

first visit data for all assessments combined. In comparing the kappa values for the assessment of chronic wounds between the PM&R and plastic surgeons, the kappa values were generally lower for the PM&R physicians, with the exception of osteomyelitis. A known paradox of kappa statistics is that kappa will be low even when agreement rate is high, when prevalence is low.²² This paradox is evident in this study for all diagnoses of repaired ulcers and for the diagnosis of cellulitis for the chronic/PM&R subgroup, in which the prevalence of the conditions was low. Their kappa values either were negative or could not be computed, even though their corresponding agreement rates tended to be high.

Accuracy

Table 4 shows the accuracy measures of sensitivity, specificity, and AUROC for the five diagnostic questions of interest. Sensitivity was slightly higher for first visits than for all visits. Specificity was generally lower for first visits, with the exception of osteomyelitis. Of note, cellulitis in the chronic/PM&R group has a sensitivity of 0%, because the telemedicine physicians diagnosed all four of the wounds with true cellulitis as “no cellulitis.” Except for osteomyelitis, the AUROC values were higher for all visit data compared with those of first visit data, suggesting an increasing level of accuracy over time. Overall, the ability of the

TABLE 4. SENSITIVITY, SPECIFICITY, AND AREA UNDER THE ROC CURVE (AUROC) MEASURING AGREEMENT BETWEEN TELEMEDICINE AND IN-PERSON ASSESSMENTS OF ALL VISIT AND FIRST VISIT^a DATA^b

		Sensitivity/Specificity		AUROC (95% CI)	
		All visits	1st visits	All visits	1st visits
Not healing	All	0.45/0.81	0.49/0.70	0.699 (0.622, 0.776)	0.663 (0.550, 0.776)
	Chronic/PM&R	0.36/0.80	0.41/0.65	0.635 (0.534, 0.737)	0.595 (0.417, 0.773)
	Chronic/plastic	0.54/0.60	0.63/0.54	0.595 (0.445, 0.745)	0.599 (0.392, 0.806)
	Repaired/plastic	0.38/0.97	0/0.88	0.860 (0.646, 1.000)	^c
	<i>p</i> -value ^d	0.13/<0.01	0.26/0.13	0.40	0.98
Necrosis	All	0.63/0.80	0.69/0.78	0.795 (0.725, 0.864)	0.756 (0.664, 0.848)
	Chronic/PM&R	0.36/0.76	0.40/0.76	0.596 (0.479, 0.713)	0.594 (0.418, 0.771)
	Chronic/plastic	0.81/0.70	0.83/0.65	0.841 (0.758, 0.925)	0.755 (0.619, 0.921)
	Repaired/plastic	0.50/0.98	^c /0.95	0.918 (0.803, 1.000)	^c
	<i>p</i> -value ^d	<0.01/<0.01	<0.01/0.08	<0.01	0.16
Cellulitis	All	0.32/0.91	0.50/0.83	0.725 (0.598, 0.852)	0.677 (0.411, 0.943)
	Chronic/PM&R	0.00/0.96	0.00/0.94	^c	^c
	Chronic/plastic	0.52/0.81	0.60/0.67	0.746 (0.619, 0.874)	0.691 (0.429, 0.953)
	Repaired/plastic	0.15/0.96	^c /0.95	0.700 (0.472, 0.928)	^c
	<i>p</i> -value ^d	0.11/<0.01	0.27/<0.01	0.76	
Osteomyelitis	All chronic	0.56/0.84	0.59/0.91	0.795 (0.721, 0.870)	0.866 (0.797, 0.935)
	Chronic/PM&R	0.64/0.71	0.68/0.83	0.740 (0.619, 0.862)	0.841 (0.716, 0.967)
	Chronic/plastic	0.37/0.92	0.38/0.94	0.779 (0.663, 0.896)	0.826 (0.713, 0.939)
	Repaired/plastic	0.11/<0.01	0.07/0.27	0.60	0.86
	<i>p</i> -value ^d	<0.01/<0.01	0.07/0.27	0.60	0.86
Not closed?	Repaired/plastic	0.59/0.99	^c /1.00	0.971 (0.923, 1.000)	^c

^aIncludes first visit after a surgical repair for initially chronic patients who were later surgically repaired.

^bTelemedicine assessments of “positively yes” or “probably yes” were categorized as “yes”; telemedicine assessments of “positively no,” “probably no,” or “not sure” were categorized as “no.”

^cToo few true negatives or true positives to obtain reliable accuracy measures.

^dThe *p*-values test for no difference in the accuracy measures of sensitivity, specificity, and AUROC among the three physician/wound type subgroups when the accuracy measures are available from all three subgroups, or between two subgroups otherwise. For all visit data, the *p*-values do not account for potential within-patient correlation.

telemedicine system to differentiate accurately between the presence and absence of a true wound condition was fair with 95% confidence intervals for the AUROC not including the value of 0.5 (no diagnostic ability), except for the first visit assessment of cellulitis.

When accuracy measures were compared across the different groups of physician and wound type combinations for first visit data, significant ($p < 0.01$) or marginally significant ($p < 0.10$) differences were observed for sensitivity in assessments of necrosis and osteomyelitis, and for specificity in assessments of necrosis and cellulitis. No significant differences were observed across groups for AUROC. In general, smaller p -values were observed for analyses of all visits, due to the larger sample size.

Comparison of agreement to baseline assessments

Before patients were enrolled for collection of the telemedicine data, two of the three pairs of participating physicians assessed a sample of patients with pressure ulcers who were different from the participants in the telemedicine study. Each physician within each pair performed the assessment independently of the other physician in the pair, and the results of the assessments were compared, to determine

a “baseline” rate of agreement within each physician pair. The participating plastic surgeons at the Ann Arbor VAMC assessed 24 patients with chronic pressure ulcers in this manner; the physical medicine and rehabilitation physicians at the Augusta VAMC assessed 23 patients.

Rates of agreement between the participating physicians and kappa statistics are presented in Table 5. As expected, baseline agreement rates and kappa statistics were generally greater than those for the telemedicine and in-person assessments; with the exception of not healing for chronic/PM&R assessments and osteomyelitis for chronic/plastic assessments. For most of the diagnoses, the observed differences in agreement rates and kappa statistics were not significantly different, with the exception of percentage agreement for cellulitis (chronic/plastic assessments) and kappa for necrosis (chronic/PM&R).

DISCUSSION

The objective of this study was to evaluate a store-and-forward telemedicine system to diagnose various wound conditions remotely. We evaluated the level of agreement between the assessments of in-person and telemedicine

TABLE 5. COMPARISON OF BASELINE AGREEMENT BETWEEN PHYSICIANS VERSUS IN-PERSON AND TELEMEDICINE AGREEMENT (FIRST VISITS)

	% Agreement (N)			Kappa (SE)		
	Baseline	In-person and telemed	p-value ^a	Baseline	In-person and telemed	p-value ^a
Not healing						
Chronic/PM&R	46.2 (13)	51.3 (39)	1.00	0.04 (0.20)	0.05 (0.15)	0.97
Chronic/plastic	70.0 (10)	58.6 (29)	0.71	0.29 (0.22)	0.16 (0.19)	0.72
Necrosis						
Chronic/PM&R	87.0 (23)	65.3 (49)	0.09	0.68 (0.21) ^b	0.17 (0.14)	0.05
Chronic/plastic	83.3 (24)	76.0 (50)	0.56	0.67 (0.20) ^b	0.49 (0.14)	0.48
Cellulitis						
Chronic/PM&R	91.3 (23)	91.8 (49)	1.00	^c	-0.03 (0.12)	^c
Chronic/plastic	100.0 (24)	66.0 (50)	<0.01	^c	0.12 (0.11)	^c
Osteomyelitis						
Chronic/PM&R	87.0 (23)	72.1 (43)	0.22	0.72 (0.21) ^b	0.42 (0.14)	0.23
Chronic/plastic	58.3 (24)	78.3 (46)	0.10	0.11 (0.20)	0.38 (0.14)	0.27

^a p -value is from Fisher's exact test, comparing baseline agreement with in-person versus telemedicine agreement.

^b $p < 0.01$ for test of kappa = 0.

^cKappa cannot be calculated due to no wounds with cellulitis.

physicians as well as the level of accuracy when telemedicine assessments were compared with in-person assessment as the gold standard assessment. The accuracy of the proposed telemedicine system to diagnose wound conditions of cellulitis, necrosis, or healing was generally fair, while the diagnosis of osteomyelitis showed a high accuracy with an AUROC of 0.87. For all conditions (diagnoses), our results showed that specificity was generally higher than sensitivity; here telemedicine assessments were dichotomized into "yes" or "no", where "positively yes" and "probably yes" were considered as the telemedicine diagnosis of "yes." Although some differences in these measures were observed across the three physician/wound subgroups, it is difficult to draw any valid conclusions from these comparisons, because of the multiple factors varying across the groups, including wound type, specialty, and years of experience of physician.

Sensitivity would be increased, and specificity decreased, if the telemedicine assessments used a more lenient cutoff in which "not sure" was also included with the "positively yes" and "probably yes" responses. Increased sensitivity is desirable to identify as many of the true positives as possible. That is, for purposes of remote diagnoses, it would be better to err on the side of overdiagnosing (i.e., diagnose the existence of a problem when one does not exist) than to underdiagnose. Decreasing specificity, however, means that there would be a greater number of patients referred for treatment or additional consultation who did not really have a condition needing treatment. The ultimate question is whether the cost of bringing in patients for possible treatment when it is not necessary outweighs the potential savings from providing more timely services to patients who truly need treatment. This should be the focus of subsequent research: to determine whether actual use of the telemedicine system improves outcomes sufficiently to justify the cost of the system.

Although our results showed that the rates of agreement between the telemedicine and in-person assessments tended to be lower than between two in-person assessments, most of these differences were not statistically significant. In addition, baseline agreement was generally

low, with kappa being lower than 0.75 for all diagnoses. These findings suggest that much of the disagreement between the in-person and telemedicine assessments may be due to differences in physician judgment. The participating physicians agreed that they probably use different criteria to respond to the diagnostic questions. To evaluate whether diagnostic concordance between in-person and telemedicine assessments is associated with technical aspects of the telemedicine system, Krupinski et al. measured correlations between dermatologists' ratings of digital image quality and diagnostic concordance and found low correlations ($r = 0.46$ between color and diagnostic concordance, and $r = 0.40$ between image sharpness and diagnostic concordance).²³ These findings further suggest that technical aspects of the telemedicine system may have little to do with observed disagreement between in-person and telemedicine assessments. Unfortunately, our study protocol did not allow us to evaluate how much of the disagreement was due to judgment differences among telemedicine evaluators or to technical failings (e.g., image quality) of the telemedicine data.

The results of this study are similar to those of Wirthlin et al., who examined the use of digital images for the assessment of vascular wounds, including postoperative incision, amputation site, necrotic/gangrenous toes, and nonhealing ulcer in a total of 38 wounds (during 45 visits) in 24 patients.⁹ Their sensitivities were higher than ours and their specificities were about the same or a little lower. Sensitivity ranged from 71% for cellulitis to 98% for wound healing and necrosis; specificity ranged from 53% for wound healing to 82% for necrosis. One reason for the higher sensitivities is that they limited their analyses to the subset of cases in which the onsite surgeons agreed more than 67% of the time. That is, they attempted to control for surgeon variation by stratifying their analyses according to the level of onsite agreement. When this was done, they observed a higher concordance between remote and onsite surgeons for cellulitis. In our study, the sample of patients used for determining the baseline agreement between the in-person physicians was not the same sample of patients used in the telemedicine portion of the study;

we therefore could not stratify our analyses in this manner. However, our results are more representative of those expected from actual use of the telemedicine system, because limiting the use of the system to only those wounds with high onsite physician agreement is not very feasible.

As with our study, the Wirthlin study observed considerable disagreement among onsite physicians: 67% for cellulitis and 80% for necrosis. (Kappa statistics were not calculated.) The authors conclude that the variability in physician judgment observed among in-person assessments limits the ability to evaluate the accuracy of digital imaging. Unfortunately, there is currently no other "gold standard" in wound assessment; the bedside examination becomes the criterion against which to compare the telemedicine assessment.

Therefore, additional research is needed to develop ways to evaluate wound conditions more objectively or, at least, more reliably, not only for improving evaluations of the accuracy of telemedicine diagnoses but also for improving the quality and consistency of care for wounds in general. Stanberry has noted⁴ that, "One of the principal dangers of telemedicine is that there may be times when it is difficult or even impossible to establish if it was a misdiagnosis by a clinician or a technical failing of the system itself that was the operative cause of harm to a telepatient." The degree of baseline disagreement observed between physicians in our study suggests that much of the source of the discrepancies between the telemedicine and in-person assessments is due to differences in physician judgment rather than to failure of the technology.

Until more research is done to determine whether discrepancies between the telemedicine and in-person assessments are due to differences in physician judgment or to problems with the telemedicine technology, we would recommend using the system to increase access to specialized wound care services if the patient does not currently receive those services (e.g., for patients residing at home or in an extended care facility, where the ulcer is managed by care givers without specialized training in wound care). Thus, the web-based telemedicine system used in this project has the poten-

tial to provide more timely assessments of wounds for patients who currently do not have access to specialized wound care. However, providers using the system must recognize its limitations (as measured by sensitivity and specificity) in achieving the same diagnoses provided by in-person assessments. The next step will be to deploy the system in several settings to improve access to specialized services, and to determine if the quality and timeliness of care do indeed improve.

ACKNOWLEDGMENTS

This work was supported by a grant from the Department of Veterans Affairs Health Services Research and Development Service, Project No. ACC 97-013-3. We are grateful to the physicians who participated in the study: Kenna Given, Jay Lucas, Kit Outlaw, Riley S. Rees, Mary D. Thomas, and Rose Trincher. We also extend our appreciation to the nurse coordinators: Ellen Clements, Mary Connors, and Cindy Hoganson.

REFERENCES

1. Smith DM, Winsemius DK, Besdine RW. Pressure sores in the elderly: can this outcome be improved? *J Gen Intern Med* **1991**;6(1):81-93.
2. Siegler EL, Lavizzo-Mourey R. Management of stage III pressure ulcers in moderately demented nursing home residents. *J Gen Intern Med* **1991**;6(6):507-13.
3. Gardner SE, Frantz RA, Specht JKP, et al. How accurate are chronic wound assessments using interactive video technology? *J Gerontol Nurs* **2001**;27(1):15-20.
4. Goldschmidt L, Hill M, Sullivan G, Perkasch I, Pejoro S, Love L. Use of telemedicine and information technology in the treatment of patients with spinal cord injury and dysfunction: improving access to specialty medical care. American Medical Informatics Association 2000 Annual Symposium, Los Angeles, CA, November 4-8, 2000.
5. Mathewson C, Adkins VK, Lenyoun MA, Schmidt AM, Jones ML. Using telemedicine in the treatment of pressure ulcers. *Ostomy Wound Manage* **1999**;45(11):58-62.
6. Vesmarovich S, Walker T, Hauber RP, Temkin A, Burns R. Use of telerehabilitation to manage pressure ulcers in persons with spinal cord injuries. *Adv Wound Care* **1999**;12(5):264-269.
7. Tomich N. Store-forward used increasingly. *U.S. Med* **1997**;33(5-6):3,24-25.

8. Scerri GV, Vassallo DJ. Initial plastic surgery experience with the first telemedicine links for the British Forces. *Br J Plast Surg* **1999**;52:294–298.
9. Wirthlin DJ, Buradagunta S, Edwards RA, Brewster D, Cambria RP, Gertler JP, LaMuraglia GM, Jordan DE, Kvedar JC, Abbott WM. Telemedicine in vascular surgery: feasibility of digital imaging for emote management of wounds. *J Vasc Surg* **1998**;27:1089–1100.
10. Burdick AE, Berman B. Teledermatology. In: Bashshur RL, Sanders JH, Shannon GW, eds. *Telemedicine: Theory and Practice*. Springfield, IL: Charles C. Thomas, **1997**;225–248.
11. Kvedar JC, Edwards RA, Menn ER, Mofid M, Gonzalez E, Dover J, Parrish J. The substitution of digital images for dermatologic physical examination. *Arch Dermatol* **1997**;133:161–167.
12. Allen MH, Burke WA, Stone D. Reliability of telemedicine for evaluating skin tumors (abstract). *Telemed J* **1997**;3(1):101.
13. Phillips CM, Burke WA, Jones BE, Laing VB, Stone D. Teledermatology at East Carolina University (abstract). *Telemed J* **1997**;3(1):106.
14. Agency for Health Care Policy and Research. Clinical Practice Guideline Number 15: Treatment of Pressure Ulcers (AHCPR Publication No. 95-0652). Rockville, MD: U.S. Department of Health and Human Services, December **1994**.
15. Lowery JC, Hamill JB, Wilkins EG, Clements E. A Web-based telemedicine system for wound assessment. *Adv Skin Wound Care* **2002**;15:165–169.
16. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* **1960**;20:37–46.
17. Diggle PJ, Liang K-Y, Zeger SL. *Analysis of longitudinal data*. Oxford, UK: Oxford University Press, **1996**.
18. Tosteson ANA, Begg CB. A general regression methodology for ROC curve estimation. *Med Decision Making* **1988**;8:204–215.
19. Weinstein MC, Fineberg HV. *Clinical decision analysis*. Philadelphia: W.B. Saunders, **1980**.
20. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* **1978**;8:283–298.
21. Tosteson ANA, Begg CB. A general regression methodology for ROC curve estimation. *Med Decision Making* **1988**;8:204–215.
22. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *J Clin Epidemiol* **1996**;49:431–434.
23. Krupinski EA, LeSueur B, Ellsworth L, Levine N, Hansen R, Silvis N, Sarantopoulos P, Hite P, Wurzel J, Weinstein R, Lopez AM. Diagnostic accuracy and image quality using a digital camera for teledermatology. *Telemed J* **1999**;5:257–263.
24. Stanberry B. Telemedicine: barriers and opportunities in the 21st century. *J Intern Med* **2000**;247:615–628.

Address reprint requests to:

Julie C. Lowery, Ph.D.

VA Center for Practice Management &

Outcomes Research

P.O. Box 130170

Ann Arbor, MI 48113-0170

E-mail: julie.lowery@med.va.gov

This article has been cited by:

1. Ulfat Shaikh , Stacey L. Cole , James P. Marcin , Thomas S. Nesbitt . 2008. Clinical Management and Patient Outcomes Among Children and Adolescents Receiving Telemedicine Consultations for Obesity. *Clinical Management and Patient Outcomes Among Children and Adolescents Receiving Telemedicine Consultations for Obesity*. *Telemedicine and e-Health* **14**:5, 434-440. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
2. A. Russell Localio, David J. Margolis, Sarah H. Kagan, Robert A. Lowe, Bruce Kinosian, Stephanie Abbuhl, William Kavesh, John H. Holmes, Althea Ruffin, Mona Baumgarten. 2006. Use of photographs for the identification of pressure ulcers in elderly hospitalized patients: validity and reliability. *Wound Repair and Regeneration* **14**:4, 506-513. [[CrossRef](#)]
3. Robert X. Murphy, Michael A. Bain, Thomas E. Wasser, Eric Wilson, Walter J. Okunski. 2006. The Reliability of Digital Imaging in the Remote Assessment of Wounds. *Annals of Plastic Surgery* **56**:4, 431-436. [[CrossRef](#)]
4. Ilias G. Maglogiannis, Kostas Karpouzis, Manolis Wallace. 2006. Image and Signal Processing for Networked E-Health Applications. *Synthesis Lectures on Biomedical Engineering* **1**:1, 1-108. [[CrossRef](#)]
5. Sophie M Jones, Paul E Banwell, Peter G Shakespeare. 2005. Telemedicine in wound healing. *International Wound Journal* **1**:4, 225-230. [[CrossRef](#)]
6. Randy S. Roth, Julie C. Lowery, Jennifer B. Hamill. 2004. Assessing Persistent Pain and Its Relation to Affective Distress, Depressive Symptoms, and Pain Catastrophizing in Patients with Chronic Wounds. *American Journal of Physical Medicine & Rehabilitation* **83**:11, 827-834. [[CrossRef](#)]