Teledermatology: An Intraobserver Diagnostic Correlation Study, Part I

MAJ Hon S. Pak, MC, USA; MAJ David Harden, MC, USA; David Cruess, PhD; LTC Mark L. Welch, MC, USA; LTC(P) Ronald Poropatich, MC, USA; and the National Capital Area Teledermatology Consortium

Many studies have been published recently on the effectiveness of teledermatology as a diagnostic tool; however, much of the data comes from live 2-way video teleconferencing consultations and very little comes from more readily available "store and forward" consultations. Moreover, most published studies compare the diagnoses of 2 different dermatologists (interobserver comparison). Given the lack of data on baseline interdermatologist diagnostic variability, the interpretation of currently available diagnostic correlation data is somewhat difficult. The objective of this study is to measure the degree of diagnostic concordance between a dermatologist seeing a patient via a teledermatology consult system and the same dermatologist seeing the same patient face-to-face in a dermatology clinic at a tertiary medical center. A random sample of 404 patients was selected from patients who had routine appointments at our dermatology clinic.

The diagnostic correlation between teledermatology and in-person consultation in this study was found to be 70% complete agreement (95% confidence level, ±4.5%), 20% partial agreement, and 10% disagreement. The authors conclude that the diagnostic agreement rate is similar to previously published studies using interobserver comparisons; therefore, teledermatology appears to be an effective method of delivering dermatologic care in the appropriate setting.

Accepted for publication February 6, 2003.

Dr. Pak is from Brooke Army Medical Center, San Antonio, Texas. Drs. Harden, Welch, and Poropatich are from Walter Reed Army Medical Center, Washington, DC. Dr. Cruess is from Uniformed Services University of Health Sciences, Bethesda, Maryland.

The authors report no conflict of interest.

The opinions expressed in this article are those of the authors and should not be construed to reflect those of the US Army or the US Department of Defense.

Reprints: MAJ Hon S. Pak, MC, USA, Department of Dermatology, Brooke Army Medical Center, 3851 Roger Brooke Dr, Bldg 3600, San Antonio, TX 78258 (e-mail: hon.pak@amedd.army.mil).

Background

To date, studies on diagnostic agreement between teledermatology and in-person evaluation have reported a 59% to 88% correlation rate.[1-11] These data are based predominantly on studies using live 2-way video teleconference consultations. Studies evaluating a more practical consultation, known as store and forward (S&F) teledermatology, have been limited in number. S&F teledermatology is the storing and asynchronously sending of patient history and clinical images via the Internet to a remote dermatologist for consultation. The few studies evaluating S&F teledermatology for diagnostic agreement, however, have been pilot trials with small numbers of patients. Whited et al[12] reported a 54% proportion of agreement among clinic-based examiners for their single most likely diagnosis and a 92% proportion of agreement when ratings included differential diagnosis. Teledermatologists had statistically similar agreement.[12]

In addition to using video teleconferencing, most published studies utilized interobserver diagnostic correlation, comparing one teledermatologist's diagnosis with a "gold standard" that usually was an in-person evaluation by a different dermatologist. However, the degree of diagnostic agreement between 2 dermatologists examining a patient in the same room has not been well established; thus, it is somewhat difficult to interpret the diagnostic correlation data reported in the literature because it was measured using interobserver agreement.

Our study is unique because it uses an intraobserver, rather than an interobserver, agreement to determine diagnostic concordance. This was done to determine if the interobserver variability represents a significant proportion of the diagnostic disagreement (noncomplete agreements) between teledermatology and in-person consultation, as reported in the literature.

Patients and Methods

Setting and Patients—The study was conducted at the Walter Reed Army Dermatology Clinic, a military tertiary medical center in Washington, DC, that is largely a referral-based outpatient clinic for military members and their beneficiaries. The study population was composed of adult patients with scheduled appointments who were randomly selected every 30 minutes. Patients less than 18 years of age, medical emergencies, or those who would be adversely affected by the additional time required to participate in the teledermatology consultation were excluded from the study. Any patients who had follow-up visits with the same dermatologist for an ongoing skin condition also were excluded from the study. In addition, any cases in which the evaluating dermatologists had prior knowledge of the patient's presenting skin condition were excluded from the study.

Study Design—This was a prospective, direct clinical comparison study conducted during a 4-month period from October 1, 1999, to January 30, 2000. Eligible patients who consented to being part of the study had a brief template-based history taken by a licensed practical nurse. The nurse then photographed images of the appropriate skin areas using an Olympus D-600L 1280×1024 (24-bit color) or a Nikon Coolpix 900 1280×960 (24-bit color) digital camera and uploaded the images and history of each patient into a computer. The images and history were viewed on a computer monitor (1024×768 resolution, 24-bit color using Netscape 4.5) by a dermatologist who used a designated form to list the rendered primary diagnosis, up to 3 differential diagnoses, and a diagnostic certainty level measured on a scale of 1 to 10, with 1 being absolutely uncertain and 10 being absolutely confident. In addition, the dermatologist listed on the form the management plan including any desired laboratory tests.

Immediately following the teledermatology consultation, the same dermatologist interviewed and examined the patient in person. At the end of the appointment, the dermatologist filled out a form similar to the one completed in the teledermatology session and listed a primary diagnosis, up to 3 differential diagnoses, category of skin disease, and any treatments and tests ordered to better define the diagnosis.

Independently, a nonparticipating dermatologist examined and compared the diagnosis and the differential diagnosis between the teledermatology and in-person consultations. A category of agreement (complete, partial, or disagreement) was then assigned; if the agreement was not complete, it was further categorized for clinical significance (mild, moderate, or severe).

Analysis of Data—Correlation of the 2 modalities was performed using a *t* test for paired data. The principal outcomes measure was the level of agreement (complete, partial, or disagreement) assigned by the independent evaluator for the 2 diagnoses received by each patient from the teledermatology and in-person evaluations. Given the sample size of 404, a 2-sided 95% confidence interval for a single proportion using the large-sample normal approximation was able to estimate to within approximately 4.5% the true complete agreement proportion. Comparison of the diagnostic certainty (confidence levels) biopsy rates was calculated using the McNemar paired χ^2 test.

Table 1.

Category and Frequency of Skin Conditions Seen in This Study Population (N=404)

Category of Disease	No.	%
Papulosquamous	105	26.0
Benign tumor	81	20.0
Premalignant/malignant	54	13.4
Eczematous	44	10.9
Pigmented lesion	34	8.4
Infectious	31	7.7
Acneform eruption	28	6.9
Other	27	6.7
Total	404	100.0

Results

Demographic Data—During a 4-month period, 404 patients who had routinely scheduled appointments were randomly selected to participate in this study. The mean age of the population was 59 years. The age range varied from 18 to 92 years. The male-to-female ratio was 1.3:1. Most of the study group (82%) were Caucasian, 13% were African American, and 5% were either Asian or Hispanic. The examiners were second- or third-year dermatology residents and staff physicians who were

Table 2.

Diagnostic Correlation Between Teledermatology and In-Person Consultation (N=404)

Type of Agreement	No.	%
Complete	283	70.0
Partial	83	20.6
Disagreement	38	9.4
Total	404	100.0

Table 3.

Clinical Significance of Partial Agreement (n=83) and Disagreement (n=38) Categories

Clinical Significance	No.	% (n=121)	Total % (N=404)
Minimal	70	57.9	17.3
Moderate	50	41.3	12.4
Severe	1	0.8	0.2

board-certified staff dermatologists with varying clinical experiences (2 to 15 years beyond dermatology residency). The category and frequency of skin conditions seen in this study population are reported in Table 1.

Diagnostic Agreement—As seen in Table 2, the diagnostic correlation between teledermatology and in-person consultation was 70% complete agreement, 20.6% partial agreement, and 9.4% disagreement. The noncomplete agreement group (partial agreement plus disagreement) comprised 30% of the total study population. When this group was further classified into levels of clinical significance, 12.6% had a clinically significant incorrect different diagnosis or different management plan in the teledermatology consultation compared with the in-person evaluation (Table 3). Table 4 lists the types of skin conditions evaluated in the study by category, with their corresponding agreement rates listed in decreasing order by the percentage of examiners in complete agreement.

Comment

Teledermatology can be best defined as the practice of dermatology using available communication and information technology. The advances in and decreasing costs of digital imaging technology, combined with the availability of the Internet, seem to put the potentials of teledermatology within reach—that is, improving patient access to dermatologic care and providing more cost-effective medicine. But how does teledermatology compare with traditional face-to-face dermatology evaluation? To date, most studies addressing this question have compared only interobserver diagnoses to determine the effectiveness of teledermatology.

Two previous studies partly used intraobserver diagnostic agreement, reporting 71% to 82% agreement[5,10]; however, the studies were done using live video teleconferencing consultations, not S&F consultations. To our knowledge, this is the first diagnostic correlation study of an S&F teledermatology consult system that utilized only an intraobserver comparison. Baseline diagnostic variability among dermatologists in a real clinical setting is not well known; therefore, it is difficult to determine the extent to which this variability affects the final diagnostic correlation between teledermatology and in-person evaluation. Our intraobserver study was an attempt to evaluate a teledermatology consult system without having to take into account this diagnostic variability among dermatologists.

In this S&F teledermatology study, 283 cases (70%) had a complete agreement between teledermatology and in-person diagnoses. These results are similar to other studies using S&F technology, which range from 61% to 88%.[3,4,13] Variations in the results of these studies are due, in part, to sample size, the definition of agreement, interobserver variability, and the exclusion of certain types of skin conditions. When complete agreements were combined with partial agreements, the overall agreement increased to 366 (90.6%). The high partial agreement rate (20.6%) most likely represents the expression of uncertainty on the part of the teledermatologist using the new technology. This increase in partial agreements among teledermatologists also was reported by Lesher et al[7] who confirmed that teledermatologists tended to give more broad differentials rather than a single diagnosis.

Because the intraobserver diagnostic correlation rates were similar to that of previous interobserver

Table 4.

Types of Skin Conditions and Their Corresponding Agreement Rates*

Category of Disease	Complete n (%)	Partial n (%)	Disagreement n (%)	Total
Infectious	26 (84)	4 (13)	1 (3)	31
Premalignant/malignant	45 (83)	5 (9)	4 (8)	54
Acneform eruption	23 (82)	4 (14)	1 (4)	28
Pigmented lesion	26 (76)	6 (18)	2 (6)	34
Benign tumor	57 (70)	17 (21)	7 (9)	81
Eczematous	29 (66)	13 (30)	2 (4)	44
Papulosquamous	62 (59)	28 (27)	15 (14)	105
Other	15 (56)	6 (22)	6 (22)	27

*A statistically significant overall difference in the distribution of levels of agreement was found across categories of disease, $P=.027$.

studies, this study can conclude that the interobserver variability does not represent a significant proportion of the difference between teledermatology and in-person evaluations (noncomplete agreement rate). Moreover, most dermatologists recognize that clinicians are faced daily with challenging cases for which a diagnosis cannot be made with certainty. Therefore, this study theorized that a significant portion of the 10% to 30% noncomplete agreement rate represents complex or difficult cases that would be difficult to diagnose even in a one-time face-to-face evaluation. Unfortunately, there is no data to conclusively support or refute this theory in the literature.

Contrary to Kvedar et al[3] who reported that no specific disease category was more or less difficult to diagnose, this study found that papulosquamous conditions, as a category, had the lowest rate of complete agreement (59%). This is consistent with the findings of Zelickson and Homan[13] who reported that eruptions were more difficult to diagnose than lesions. Loane and colleagues[10] also reported that eczematous conditions accounted for more than one third of all the inappropriate management decisions. Although it is unclear why, the inability to perform in-office tests (such as a potassium hydroxide) and to palpate the lesion(s) also may contribute to the lower diagnostic correlation and higher uncertainty.

Acknowledgement—The authors would like to thank Mrs. Linda McKnight for the organization of and data collection for this study; the staff and residents of the Dermatology Service, Walter Reed Army Medical Center, for their hard work and support of this study; and John D. Whited, MD, for his critical review of and recommendations for this manuscript.

REFERENCES

1. Harrison PV, Kirby B, Dickinson Y, et al. Teledermatology—high technology or not? *J Telemed Telecare*. 1998;4(suppl 1):31-32.
2. Phillips CM, Burke WA, Allen MH, et al. Reliability of telemedicine in evaluating skin tumors. *Telemed J*. 1998;4:5-9.
3. Kvedar JC, Edwards RA, Menn ER, et al. The substitution of digital images for dermatologic physical examination. *Arch Dermatol*. 1997;133:161-167.
4. Whited JD, Mills BJ, Hall RP, et al. A pilot trial of digital imaging in skin cancer. *J Telemed Telecare*. 1998;4:108-112.
5. Oakley AM, Astwood DR, Loane M, et al. Diagnostic accuracy of teledermatology: results of a preliminary study in New Zealand. *N Z Med J*. 1997;110:51-53.
6. Phillips CM, Burke WA, Shechter A, et al. Reliability of dermatology teleconsultations with the use of teleconferencing technology. *J Am Acad Dermatol*. 1997;37:398-402.

7. Lesher JL, Davis LS, Gourdin FW, et al. Telemedicine evaluation of cutaneous diseases: a blinded comparative study. *J Am Acad Dermatol.* 1998;38:27-31.
8. Lowitt MH, Kessler II, Kauffman L, et al. Teledermatology and in-person examinations. *Arch Dermatol.* 1998;134:471-476.
9. Gilmour E, Campbell SM, Loane MA, et al. Comparison of teleconsultations and face-to-face consultations: preliminary results of a United Kingdom multicentre teledermatology study. *Br J Dermatol.* 1998;139:81-87.
10. Loane MA, Corbett R, Bloomer SE, et al. Diagnostic accuracy and clinical management by realtime teledermatology. results from the Northern Ireland arms of the UK Multicentre Teledermatology Trial. *J Telemed Telecare*. 1998;4:95-100.
11. Loane MA, Gore HE, Corbett R, et al. Preliminary results from the Northern Ireland arms of the UK Multicentre Teledermatology Trial: effect of camera performance on diagnostic accuracy. *J Telemed Telecare*. 1997;3(suppl 1):73-75.
12. Whited JD, Russell HP, Simel DL, et al. Reliability and accuracy of dermatologists' clinic-based and digital image consultations. *J Am Acad Dermatol.* 1999;41:693-702.
13. Zelickson BD, Homan L. Teledermatology in the nursing home. *Arch Dermatol.* 1997;133:171-174.