



ACTAS Dermo-Sifiliográficas

Full English text available at
www.actasdermo.org



ORIGINAL ARTICLE

Interobserver Reliability of Store-and-Forward Teledermatology in a Clinical Practice Setting[☆]



G. Romero Aguilera,^{*} P. Cortina de la Calle, E. Vera Iglesias,
P. Sánchez Caminero, M. García Arpa, J.A. Garrido Martín

Servicio de Dermatología, Hospital General Universitario de Ciudad Real, Ciudad Real, Spain

Received 20 August 2013; accepted 11 December 2013

Available online 14 June 2014

KEYWORDS

Telemedicine;
Teleconsultation;
Teledermatology;
Store-and-forward
systems;
Reliability

Abstract

Introduction and objectives: Although many studies have evaluated the diagnostic reliability of store-and-forward (SF) teledermatology, the reliability of the technique for the diagnosis of general skin conditions in a clinical practice setting has never been demonstrated. We evaluated the reliability of SF teledermatology in clinical practice by analyzing the diagnostic agreement achieved in a subgroup of patients from the DERMATEL-2 study.

Material and methods: Patients referred from primary care settings were randomized to 3 groups: SF, a combination of videoconferencing and SF technology (VC-SF), and a control group. This article focuses on the SF group. Clinical data were recorded and photographs taken by primary care physicians, who forwarded the data digitally. Each SF consultation package was assessed by 3 dermatologists (D1, D2, D3). Subsequently all the patients were assessed by a single dermatologist (D1) in a face-to-face (FF) consultation. Finally, 2 other dermatologists (D4, D5) assessed the agreement between the diagnoses obtained by SF and FF.

Results: In total, 457 patients (200 males and 257 females) aged between 2 months and 86 years were randomized (192 to SF, 176 to VC-SF, and 89 to the control group). The diagnostic categories were as follows: tumors (49.4%), inflammatory (25.7%), adnexal (11%), infectious (9.4%) and other processes (4.4%). Since 170 patients had SF consultations deemed valid for analysis, the study included a total of 510 SF assessments. Most of the images and clinical records were of high quality (71.2% and 91.2% respectively), and diagnostic confidence was high in 81.4% of the cases studied.

In 58.4% of cases the condition was managed exclusively by teledermatology. Levels of complete and aggregate interobserver agreement between SF and FF evaluators were 0.72 and 0.90, respectively, for diagnosis and 0.61 and 0.80 for treatment. Diagnostic agreement correlated with the image quality ($P < .001$), diagnostic confidence ($P < .001$), felt need for conventional consultation ($P < .001$), and the quality of the clinical record ($P = .013$).

[☆] Please cite this article as: Romero Aguilera G, Cortina de la Calle P, Vera Iglesias E, Sánchez Caminero P, García Arpa M, Garrido Martín J. Fiabilidad de la teledermatología de almacenamiento en un escenario real. Actas Dermosifiliogr. 2014;105:605–613.

^{*} Corresponding author.

E-mail address: gromeroa@gmail.com (G. Romero Aguilera).

PALABRAS CLAVE

Telemedicina;
Teleconsulta;
Teledermatología;
Sistemas de
almacenamiento;
Fiabilidad

Conclusion: The interobserver reliability of SF diagnosis in clinical practice is good. Dermatologists are able to predict errors in diagnosis by analyzing their own diagnostic confidence and evaluating the quality of the images.

© 2013 Elsevier España, S.L. and AEDV. All rights reserved.

Fiabilidad de la teledermatología de almacenamiento en un escenario real

Resumen

Introducción y objetivos: Aunque existen múltiples estudios de fiabilidad diagnóstica en teledermatología de almacenamiento (TDA), aún no se ha demostrado una fiabilidad elevada para enfermedad general cutánea en un escenario real. DERMATEL-2 fue un estudio aleatorizado de concordancia diagnóstica en TDA en condiciones de práctica clínica.

Material y métodos: Pacientes remitidos desde atención primaria fueron aleatorizados en 3 grupos: TDA; híbrida videoconferencia-almacenamiento (VC-TDA) y grupo control (GC). Este artículo se centra en el grupo de almacenamiento. Médicos de atención primaria tomaron datos y fotografías clínicas remitiéndolas a distancia. Cada consulta de TDA fue evaluada por 3 dermatólogos diferentes (D1,D2,D3). Todos los pacientes fueron finalmente vistos por el mismo dermatólogo (D1) en la consulta presencial (CP). Dos dermatólogos adicionales (D4, D5) evaluaron las concordancias TDA-CP.

Resultados: Se aleatorizaron un total de 457 pacientes 4-4-2: 192 TDA, 176 VC-TDA y 89 GC; 200 varones y 257 mujeres, 0-86 años. Se incluyó enfermedad tumoral (49,4%), inflamatoria (25,7%), anéxica (11%), infecciosa (9,4%) y otros (4,4%). Hubo 170 pacientes de TDA válidos para el análisis, rindiendo 510 teleconsultas-TDA. La imagen (71,2%), la historia clínica (91,2%) y la confianza diagnóstica (81,4%) fueron de calidad alta. En el 58,4% fue posible el manejo exclusivo *on-line*. Los acuerdos interobservador (completo/agregado) TDA-CP fueron 0,72/0,90 para el diagnóstico y 0,61/0,80 en el tratamiento. El acuerdo diagnóstico se correlacionó con la calidad de la imagen ($p < 0,001$), la confianza diagnóstica ($p < 0,001$), la necesidad de consulta presencial ($p < 0,001$) y la calidad de la historia clínica ($p = 0,013$).

Conclusión: La fiabilidad diagnóstica de la TDA en condiciones de práctica clínica es elevada. Los dermatólogos pueden predecir errores diagnósticos analizando su confianza diagnóstica y la calidad de las fotografías.

© 2013 Elsevier España, S.L. y AEDV. Todos los derechos reservados.

Introduction

A growing number of patients, physicians, and institutions are using telemedicine services,¹ and the world telemedicine market is expected to grow from US\$9.8 billion in 2010 to some \$US27.3 billion in 2016.² Dermatology appears to be an ideal specialty for telemedicine.³ Indeed, it is the clinical specialty that currently accounts for the largest number of telemedicine studies.¹ Teledermatology can be carried out in real time using videoconferencing systems (VC) or asynchronously using store and forward (SF) systems, a modality in which clinical images are obtained by the physician, stored, and then forwarded to the specialist for later assessment. VC requires greater coordination and consumes more time and resources.^{4,5} It can also be more costly than conventional face-to-face (FF) consultation.⁶ While VC allows the specialist to interact directly with the patient, the reliability of the method (as measured by diagnostic agreement) can be similar to that of SF teleconsultation.^{4,7,8} Thus, medical professionals and healthcare providers increasingly consider SF to be the most cost effective and convenient system.⁹

The teleconsultation is a complex intervention that depends on multiple interactions and is influenced by local administrative and organizational variations. As a result, it is difficult in telemedicine to conduct studies that provide a high level of evidence, such as randomized clinical trials.

A systematic review concluded that teledermatology is far from being a mature technology,¹⁰ and that more randomized trials involving real patients are needed to provide high quality evidence.

To date, most of the research in teledermatology has assessed reliability on the basis of concurrence in diagnosis.¹¹ Most of these studies distinguish between complete agreement (concurrence on a single diagnosis) and aggregate agreement (concurrence on at least one diagnosis when a differential diagnosis is proposed) (Table 1). Although it can be high, diagnostic reliability does vary considerably, with the percentage of complete agreement ranging from 47% to 90% and for aggregate agreement from 60% to 99%.^{1,11} While many authors have studied the diagnostic reliability of SF, none of the studies have been randomized controlled trials and almost all were conducted in an experimental rather than a clinical setting.¹⁰

In routine clinical practice, it is the primary care physician or other health professional who, in a remote location, typically a primary care center, sees the patient, records the clinical history, and takes the photographs. The clinical data is then forwarded through an intranet or the Internet to the hospital-based dermatologist for assessment. However, in nearly all of the published studies on the reliability of SF, the clinical history and the photographs were taken on site in the hospital itself (not in a remote location) and they were not taken by primary care physicians, but rather

Table 1 Forms and Scales of Agreement Used in the DER-MATEL Study.

A) Data collected on the form completed by primary care physicians during the initial consultation
1. Personal and family medical history
2. Reason for consultation
3. Duration of the clinical condition
4. Site of the lesions
5. Symptoms (pruritus, pain, etc.)
6. Diagnosis
7. Treatment
B) Data collected on the form completed by the dermatologists
1. Diagnosis: defined as either the most likely single diagnosis or as a differential diagnosis. ICD-9 coding modified for dermatology was used
2. Recommended treatment
3. Recommended diagnostic tests including biopsy, microbiology, blood and urine tests, and other explorations
4. Need for follow-up (yes or no)
5. Quality of the digital photos received (good or poor)
6. Quality of the medical history supplied by the primary care physical (good or poor)
7. Confidence of the dermatologist in his or her telediagnosis, categorized on a 5-item Likert scale (see text)
8. Need for a face-to-face consultation after teleconsultation, categorized as follows: unnecessary, required for diagnosis, required for management, or required for diagnosis and management
C) Scales of agreement
<i>I. Agreement on diagnosis</i>
1. Complete agreement (CA): a single diagnosis was recorded and the specialists concurred
2. Partial agreement (PA) at least one differential diagnosis was recorded and the specialists concurred on at least one possible diagnosis
3. Aggregate agreement: the sum of CA and PA
4. Disagreement: no diagnostic concordance on either a single or differential diagnosis
<i>II. Agreement on treatment</i>
1. CA: identical treatment prescribed following both types of consultation
2. PA: small differences between the consultation modalities in the treatment prescribed, including acceptable therapeutic variations
3. Aggregate agreement: the sum of CA and PA
4. Disagreement: different treatments prescribed (unacceptable variation in therapy)

by experts, whether dermatologists, dermatology residents, professional photographers, or research assistants.^{11,12} This bias could artificially increase the apparent reliability of SF in these studies.

In fact there are no studies that demonstrate the reliability of SF in general skin disease in routine clinical practice. This was the main objective of the present study: to assess interobserver reliability in SF when the patients were assessed in a remote location under the conditions of routine clinical practice. A secondary objective was to

investigate whether teledermatologists were able to predict their own diagnostic errors under these conditions.

Material and Methods

Study Design

DERMATEL was a prospective, randomized, controlled, experimental study of diagnostic agreement. The study design was based on multiple assessments of each patient. The only inclusion criterion was that the patient be referred from primary care to a specialist for a skin condition that had never been assessed by a dermatologist. Patients who agreed to participate signed an informed consent form approved by the hospital's research committee.

A total of 457 patients referred from 6 primary care centers between June 2004 and December 2005 were included in the study. Patients were randomized (4:4:2) to 3 study groups by a computerized system that centralized all the data via the Internet: 192 patients were randomized to the SF group, 176 to the VC-SF hybrid group, and 89 to the control group. The results for the VC group have already been published,⁴ and this article deals with the results for the SF group.

Five dermatologists, 18 primary care physicians (12 general practitioners and 6 pediatricians) participated in the study. All of the participating physicians received training in the use of the software, and the primary care physicians attended photography workshops. The trainers emphasized the need for at least 2 to 3 photographs in each case, including general shots identifying the site and the distribution of the lesions in addition to more detailed close-up images. Participants were instructed on the need for proper room lighting and the obligatory use of flash for all shots. The dermatologists were staff physicians with an average of 8 years of clinical experience, and only one of them had experience in teledermatology (3 years) at the time of the study.

Once the patient was randomized to a group, the primary care physicians took the clinical photographs and completed the standard consultation form. This information was then forwarded to the hospital website. Each of the patients in the SF group was assessed on-line by 3 dermatologists (D1, D2, and D3) on the basis of this teleconsultation package. Two days after each teleconsultation had been assessed, 1 of the 3 dermatologists (D1) assessed the same patient in an FF consultation, considered to be the gold standard modality. The patients in the control group were only assessed in an in-person consultation. The management of the hybrid VC-SF group formed part of another study, the results of which have already been published.⁴

Technical Aspects and Data Collection

The software for managing the remote consultations (DERMARED) was designed specifically for the purpose and is a web-based application with a centralized database hosted on the Castilla la Mancha Health Services (SESCAM) intranet. Clinicians took the photographs with midrange digital cameras using JPEG compression (Sony Cybershot DSC-F717 or Nikon Coolpix 4500).

Four forms were completed for each patient: demographic data, clinical data collected by the primary care

physician, data supplied by the dermatologist, and a form for comparing consultations according to predefined scales of agreement (Table 1).

Data Analysis

The 3 dermatologists rated the degree of confidence they had in their own diagnosis in each case on an appropriate Likert scale. On this scale, the rater chose from a continuous series of alternatives ranging from maximum to minimum, classifying their confidence in the diagnosis as very high, high, medium, low, or very low.

The data were processed using the SPSS statistical package (SPSS Inc). Two other dermatologists (D4 and D5) independently evaluated the concordance between the data collected in the remote and in-person consultations for each patient. An analysis was performed using the χ^2 test for 2 × 2 tables. To ascertain whether the teledermatologists were able to predict their own diagnostic errors, the percentages of agreement were stratified by their rating of diagnostic confidence, image quality, the quality of the clinical history provided by the primary care physician, and their opinion on the need for an in-person consultation to assess the patient.

Results

Description of the Sample

A total of 457 patients were included in the study: 200 men (44%) and 257 women (56%); the mean age was 36 years (range: 2 mo–86 y). The distribution by diagnostic category was as follows: tumors (49.4%), inflammatory (25.7%), adnexal (11%), infectious (9.4%), and other processes (4.4%).

There were no significant differences between the SF, VC-SF, and control groups in terms of age (ANOVA, $P = .057$), sex ($\chi^2 = 2.13$, $P = .334$), or diagnostic category ($\chi^2 = 14.538$ and $P = .268$).

In total, 11.4% of patients in the SF group and 10.2% in the VC-SF group did not attend the FF consultation; the difference was not significant ($P = .417$). In total, 170 of the 192 patients randomized to the SF group were valid for analysis. The present article discusses the results for the SF group. Since each patient was assessed by all 3 dermatologists, 510 SF teleconsultations were undertaken in total.

Diagnostic Agreement

Table 2 shows the diagnostic agreement between dermatologists and modalities (SF vs FF) as well as the intramodal concordance (SF vs SF). In the interobserver comparison (D2 vs D1 and D3 vs D1), complete and aggregate agreement between the outcomes of SF and FF consultations were high for both diagnosis (0.72 and 0.90, respectively) and treatment (0.61 and 0.80, respectively). Intraobserver diagnostic agreement (D1 vs D1) was significantly higher than interobserver agreement ($P = .0001$).

Factors Modifying Agreement

In the group of patients who had SF teleconsultations, the dermatologists rated the quality of the photographs as poor

in 18% of cases and the quality of clinical histories as poor in 8.8%. No in-person consultation was deemed necessary in 58% of the patients assessed (Table 3). The dermatologists' confidence in their own diagnosis was high in 81% of cases. There were differences between the dermatologists with respect to these 4 factors ($P < .001$). The youngest dermatologist (D2) was more demanding with respect to the quality of the photographs and the least demanding regarding the quality of the clinical history. D2 was also the most uncertain about the diagnosis and about exclusively on-line management of the case. The dermatologist with prior experience in teledermatology (D1) was more confident about diagnosis and recommended an FF consultation in the smallest number of cases.

There was a statistically significant relationship between diagnostic agreement, diagnostic confidence, the perceived need for an in-person consultation, the image quality rating, and—to a lesser degree—the quality of the clinical history (Table 4). Complete agreement on diagnosis was over 78% in the group in which an in-person consultation was not deemed necessary. However, this percentage fell to 43% in the group of patients for whom an in-person consultation was considered essential for a valid diagnosis ($\chi^2 = 64.301$; $P < .001$).

Complete agreement was higher for acne (90.5%) and infections (80.4%), including common warts. The difference was significant for acne ($P = .03$), but not for infections ($P = .083$). CA was similar for tumors (69%) and inflammatory conditions (65%), with a P value of .4421.

Analysis of Agreements and Errors

With respect to serious errors, only 1 of the 3 teledermatologists correctly diagnosed a tongue carcinoma in the teleconsultation. However, all 3 teleconsultants recommended in-person consultation and biopsy. The photographs in that particular case were rated as poor quality by all 3 dermatologists, although they were sufficiently clear to allow all 3 specialists to identify the elementary lesion as an ulcer.

By contrast, on the basis of the teleconsultation all 3 specialists correctly diagnosed the 9 cases of nonmelanoma skin cancer in the study: 6 basal cell carcinomas, 1 squamous cell carcinoma of the skin, 1 squamous cell carcinoma of the lip, and 1 keratoacanthoma. None of the patients in the sample had melanoma. Only 3 of the 170 cases were not diagnosed correctly on the basis of the teleconsultation by any of the 3 dermatologists: a case of pityriasis rosea, a case of nevus depigmentosus, and a case of juvenile xanthogranuloma. In a further 4 cases, there were marked discrepancies (Table 5). By contrast, in 137 of the 170 patients (80.5%) aggregate diagnostic agreement between the 4 assessments—the 3 teleconsultations and the in-person consultation—was perfect (Table 6).

Agreement on Diagnostic Tests Ordered, Follow-Up, and Treatment

The level of agreement between the dermatologists in SF was very high, both in the diagnostic tests they ordered (in which they concurred in >80% of cases) as well as in the decision on the need for patient follow-up (agreement

Table 2 Diagnostic Agreement: Intermodality (SF vs FF) and Intramodality Between the 3 Teledermatologists (SF vs SF).^a

Diagnostic Agreement	Intermodality: SF vs FF			
	SF D2-FF D1 ^b n = 170	SF D3-FF D1 n = 170	SF D1-FF D1 n = 170	SF-FF × 3 (95% CI)
Complete agreement	72.9	72.4	88.2	77.8 (74.1-81.4)
Partial agreement	15.3	20.0	7.6	14.3 (11.2-17.3)
Disagreement	11.8	7.6	4.1	7.8 (5.4-10.1)
Diagnostic Agreement	Intramodality SF vs SF (Reliability)			
	SF D1-SF D2 n = 170	SF D1-SF D3 n = 170	SF D2-SF D3 n = 170	SF-SF × 3 (95% CI)
Complete agreement	69.4	72.9	67.1	69.8 (65.8-73.7)
Partial agreement	15.3	15.9	19.4	16.9 (13.6-20.1)
Disagreement	15.3	11.2	13.5	13.3 (10.3-16.2)

Abbreviations: FF, face-to-face in-person consultation; SF, store and forward teleconsultation.

^a All values are expressed as percentages except where otherwise indicated.

^b The 3 participating dermatologists are designated D1, D2, and D3.

Table 3 Need for Face-to-Face Consultation (FF) in the 510 Store and Forward (SF) Teleconsultations.^a

Need for FF consultation	SF-D1	SF-D2	SF-D3	Total
For diagnosis	23 (13.5%)	30 (17.6%)	36 (21.2%)	89 (16.3%)
For management	4 (2.4%)	40 (23.5%)	34 (20.0%)	78 (13.5%)
For diagnosis and management	14 (8.2%)	32 (18.8%)	16 (9.4%)	62 (11.8%)
Not required	129 (75.9%)	68 (40.0%)	84 (49.4%)	281 (58.4%)
Total	170 (100%)	170 (100%)	170 (100.0%)	510 (100.0%)

^a The 3 participating dermatologists are designated D1, D2, and D3

in > 70% of cases). The treatment prescribed was identical in 51.3% of cases and similar in a further 20.8% (Table 7). Thus, there was clearly greater disagreement concerning treatment (27.8%) than diagnosis (13.3%).

Discussion

In a routine clinical practice setting, the rate of complete and aggregate diagnostic agreement was 75% and 90%, respectively, in the intermodality comparison (SF vs FF) and 70% and 85%, respectively, in the intramodality comparison (SF vs SF). These findings are comparable to the highest concordance rates published in more than 40 studies on the

reliability of SF and show that SF is very reliable in the clinical setting.⁵

While the accumulated evidence on this topic is of high quality, it may be weakened by certain types of bias that we have attempted to avoid our study.^{1,13} DERMATEL was a randomized prospective trial that included interobserver comparisons. The sample was representative of the general population as it included all types of disorders (not just tumors) and had an appropriate age and sex distribution. The sample used for SF comparisons was large—510 teleconsultations (170 × 3)—and, most importantly, the study took place in a routine clinical setting: 18 primary care physicians took photographs and recorded the clinical data remotely in

Table 4 Diagnostic Agreement in Store and Forward Teleconsultations by Potential Modifying Factors.

Modifying Factors	Rating	Diagnostic Agreement, %			Analysis	
		Complete Agreement	Partial Agreement	Disagreement	χ ²	P Value
Diagnostic Confidence	Low	47.2	18.1	34.7	34.960	P < .001
	High	73.5	16.7	9.8		
Need for face-to-face consultation	Yes	43.1	25.2	31.7	64.301	P < .001
	No	78.3	14.2	7.5		
Image quality	Poor	57.6	17.4	25.0	22.029	P < .001
	Good	74.1	16.7	9.3		
Quality of medical history	Poor	46.7	33.3	20.0	8.635	P = .013
	Good	71.3	15.8	12.9		

Table 5 Illustrative Instances of Discrepancies Between the 4 Assessments: 3 Teleconsultations (D1, D2, D3) and 1 Face-to-Face Consultation (D1).

Case No	FF D1	SF-D1	SF-D2	SF-D3	AP	FF-R	DC	PQ
Case 1	Pilomatricoma	Dermatofibroma vs epidermal cyst	Epidermal Cyst	Dermatofibroma vs angioleiomyoma	Juvenile xanthogranuloma	Yes	Low	Poor
Case 2	Pityriasis rosea	Seborrheic dermatitis	Acute eczema vs seborrheic dermatitis	Seborrheic dermatitis	NA	Yes	Low	Poor
Case 3	Nevus depigmentosus	Lichen sclerosus et atrophicus vs vitiligo	Vitiligo vs lichen sclerosus et atrophicus	Vitiligo	NA	Yes	Low	Poor
Case 4	Granuloma annulare vs common wart	Mosaic warts	Common wart	Granuloma annulare	Granuloma annulare	Yes	Low	Medium
Case 5	Carcinoma of the tongue	Carcinoma of the tongue	Major aphthous ulcer vs traumatic ulcer	Eosinophilic ulcer	Carcinoma of the tongue	Yes	Low	Poor
Case 6	Alopecia in discoid lupus	Alopecia areata	Scarring alopecia	Alopecia	Discoid lupus	Yes	Low	Poor
Case 7	Drug-induced LE	Drug-induced LE	Subacute LE vs drug induced LE	Photodermatitis	Lupus erythematosus	Yes	High	Good

Abbreviations: AP, anatomical pathology; FF, face-to-face consultation; FF-R, face-to-face consultation required; DC, diagnostic confidence; LE, lupus erythematosus; NA, not applicable; PQ, photographic quality; SF, store and forward teleconsultation.

health care centers. The teleconsultations were then transmitted digitally over the Internet. Thus, the characteristics of the study guarantee a high degree of internal and external validity.¹

Most earlier studies have assessed the reliability of SF under ideal "laboratory" conditions, unlike those of routine workflow in a normal clinical setting.^{1,10,11} In the study by Krupinski et al.¹⁴ the photographs were taken by a medical student under the direction of a dermatologist, and in the study by Lyon et al.¹⁵ the photographs were taken by a resident in dermatology. In one study,¹⁶ the photographs were taken by a professional photographer, and in others by a research assistant.^{17,18} Finally, in a number of other studies, photographs were taken by a nurse trained in clinical photography.^{7,19-21} In most of these studies, the images and clinical data were obtained on site in the hospital where

they were later assessed by the 'tele' dermatologist.^{7,15-21} This procedure made it easier to carry out the study but may have introduced bias.

It has been observed that the results are poorer when several different primary care physicians are responsible for taking the photographs and collecting the clinical data and when the teleconsultations are sent from a remote location.^{12,22,23} Only 2 earlier SF studies were similar to DER-MATEL in that they assessed all types of skin disease (tumors and other processes) and emulated the conditions of routine clinical practice.^{12,22} No valid conclusions can be drawn from one of these studies due to flaws in the study design.²² The other produced the worst results of the 41 studies on SF in the literature, reporting complete agreement of 54% and aggregate agreement of 63%.^{12,13} The authors of that study attributed the poor results obtained to their use of real

Table 6 Cases With Perfect Aggregate Diagnostic Concordance Between the 4 Assessments (3 Teleconsultations and 1 In-Person Consultation) Achieved in 137 out of 170 cases (80.5%).

Benign tumors (56): melanocytic nevus (17 acquired, 4 congenital, 1 halo), seborrheic keratosis (14), epidermal cysts (4), dermatofibroma (3), capillary angioma (2), sebaceous nevus (2), irritated seborrheic keratosis (1), acrochordon (1), solar lentigo (1), xanthelasma (1), digital mucous cyst (1), juvenile xanthogranuloma (1), verrucous epidermal nevus (1), solitary mastocytoma (1), superficial actinic porokeratosis (1)

Inflammatory Conditions (28): atopic dermatitis (5), psoriasis vulgaris (5), contact dermatitis (3), stasis dermatitis (2), pityriasis amiantacea (2), pigmented purpura (2), pityriasis alba (1), acute urticaria (1), insect bite (1), granuloma annulare (1), pruritus sine materia (1), morphea (1), lichen planus (1), lichen sclerosus et atrophicus (1), bullous pemphigoid (1)

Infectious Conditions (19): common warts (8), molluscum contagiosum (4), genital warts (2), pityriasis versicolor (2), plantar warts (1), herpes zoster (1), candidal intertrigo (1)

Malignant Tumors (15): Basal cell carcinoma (6), actinic keratosis (5), hypertrophic actinic keratosis (1), keratoacanthoma (1), squamous cell carcinoma of the skin (1), squamous cell carcinoma of the lip (1).

Adnexal (15): Acne vulgaris (6), androgenic alopecia (4), rosacea (2), acne conglobata (1), onychodystrophy (1), subungual hematoma (1)

Others (4): melasma (1), hypomelanosis of Ito (1), stretch marks (1), telangiectasias (1)

Table 7 Intermodality Treatment Agreement (SF vs FF) and Intramodality Treatment Agreement (SF vs SF)^a

Treatment Agreement	Intermodality: SF vs FF (Validity)			
	SF D2-FF D1 ^b n = 170	SF D3-FF D1 n = 170	SF D1-FF D1 n = 170	SF-FF × 3 (95% CI)
Complete agreement	57.1	65.9	85.3	69.4 (74.1-81.4)
Partial agreement	20.0	17.6	7.1	14.9 (11.2-17.3)
Disagreement	22.9	16.5	7.6	15.6 (5.4-10.1)
Treatment Agreement	Intramodality SF vs SF (Reliability)			
	SF D1-SF D2 n = 170	SF D1-SF D3 n = 170	SF D2-SF D3 n = 170	SF-SF × 3
(95% CI)				
Complete agreement	44.1	52.4	57.6	51.3 (65.8-73.7)
Partial agreement	17.1	25.9	19.4	20.8 (13.6-20.1)
Disagreement	38.8	21.8	22.9	27.8 (10.3-16.2)

Abbreviations: FF, face-to-face in-person consultation; SF, store and forward teleconsultation.

^a All values are expressed as percentages except where otherwise indicated.

^b The 3 participating dermatologists are designated D1, D2, and D3

conditions rather than the ideal conditions created in the other studies.¹² Some studies focusing exclusively on skin tumors have achieved good reliability in a clinical setting, an outcome that supports our findings.²⁴ An important objective of the present study was to demonstrate the diagnostic reliability of SF in a clinical setting and not only for tumors but also for skin disease in general.

The design of the DERMATEL study included the use of a predefined medical history form containing a clear set of items, some of which were mandatory. Moreover, the protocol specified that the primary care physicians had to receive training in digital photography. Data storage and processing was centralized in the hospital. All of these characteristics may explain why the results of our study tend to be better on average than those reported in the literature.

Intraobserver diagnostic agreement was very high in our study, with values comparable to those reported by the study in the literature reporting the best results for reliability.²⁵ In addition, our findings relating to reliability (SF vs SF intramodality agreement) are similar to those published previously.^{7,14,16,17}

In DERMATEL, diagnostic agreement correlated with the teledermatologists' confidence in their own diagnosis and was associated with the recommendation that an in-person consultation was needed and, to a lesser extent, with the quality of the images and clinical data in the teleconsultation (Table 2). It would appear, therefore, that the teledermatologist can evaluate the likelihood of an error in diagnosis and avoid this outcome by recommending a conventional in-person consultation. Although the correlation between diagnostic agreement and diagnostic confidence has been studied extensively in the literature,^{7,14,18} the present study was the first to analyze the correlation between diagnostic agreement and the decision to recommend an in-person consultation.⁴

Our statistical analysis showed with a high degree of statistical significance that the teledermatologist can predict and therefore avoid errors in diagnosis ($P < .001$ and $\chi^2 > 34.9$) (Table 4). If the teledermatologist considers the

quality of the images and clinical history to be adequate and has a high degree of confidence in his or her own diagnosis, the likelihood of an error is very low. In cases in which the quality of the images or of the clinical history is rated as inadequate or diagnostic confidence is low, the dermatologist can prevent any error in diagnosis by recommending further assessment in an FF consultation.

The results of DERMATEL also showed that the dermatologist, as an expert morphologist, can easily make a diagnosis when the case is typical and the image provided is of good quality. Logically, if the case is atypical or complex (e.g., autoimmune disease, severe psoriasis, or contact dermatitis), the dermatologist must see the patient in person at the clinic.

The error analysis showed that very diverse disorders (Table 6) can be diagnosed with a high degree of reliability. The cases assessed included a wide variety of disorders: common conditions such as common warts and melanocytic nevae; rare diseases, such as superficial actinic porokeratosis and solitary mastocytoma; and potentially serious disorders, including nonmelanoma skin cancer and bullous disorders. The type of cases that gave rise to disagreements illustrate some of the limitations of SF (Table 5). As has been suggested previously, lesions located in the mouth (case 5) and on the scalp (case 6) are more difficult to diagnose in a teleconsultation because they are more difficult to photograph.²⁶ Similarly, the diagnosis of nodules (case 1) poses problems because of the importance of palpation in their assessment. In all the cases where there were important differences of opinion, the photographs were of poor quality and the dermatologists recommended in-person consultation.

Many skin conditions can be managed on-line, thus avoiding the need for in-person consultations. In other cases, despite a correct diagnosis, an FF consultation is necessary for the management of the condition, whether because of the need for diagnostic tests or therapeutic procedures (Table 3). SF can only eliminate the need for an in-person

consultation in 50% of cases.²⁷ Online presurgical planning may eliminate the need for a further 15% of live consultations when patients can go directly to a surgical intervention following a teleconsultation.²⁸

The DERMATEL study was completed in 2007. Since then our group has developed a software package (DERCAM) that integrates digital photographic images into the radiology picture archiving and communication system (PACS) and facilitates SF teledermatology. Since this system was introduced in our hospital in 2009, 4 dermatologists have seen over 2000 patients and an FF consultation has been avoided in 53% of these cases. During this period we have continued to assess both the quality of the photographs and the dermatologist's confidence in his or her diagnosis. The results obtained in a real clinical setting have been comparable to the experimental results of the DERMATEL study: high diagnostic confidence in 74% of cases and good photographic quality in 81%.

A systematic review of the literature revealed the need for more quality research in teledermatology to provide quality evidence supporting the use of this technology in routine clinical practice.¹⁰ We believe that our study and other recent studies provide evidence to support the usefulness of teledermatology in routine practice.^{29–32} One randomized clinical trial showed that SF can be a cost-saving strategy for providing dermatological care,³⁰ producing clinical outcomes similar to those obtained with conventional consultations.³¹ A recent study demonstrated the feasibility of SF on a large scale²⁹: 1820 general practitioners and 166 dermatologists performed more than 37 000 SF consultations between 2007 and 2010. There was no need for an FF consultation in 74% of these cases. The authors of that study concluded that SF should be considered a possible pathway of referral from primary to secondary care and we agree with this conclusion.

The diagnostic validity of SF for skin tumors compared to the gold standard of histologic study is very high, but that of FF consultation in such cases is even higher.⁵ The DERMATEL study has shown that teledermatologists can predict and prevent errors by recommending a live consultation when they have any doubts. For example, an analysis of the literature demonstrates the greater diagnostic ability of FF consultation vs teledermatology in the case of pigmented lesions,³³ but it has recently been shown that the routine use of SF in a practice setting can enhance early diagnosis of malignant melanoma because the teledermatology procedure involves more active collaboration on the part of the primary care physicians.³² Teledermatology can also offer other advantages over in-person consultation: greater accessibility, shorter time frame, lower costs, reduced patient travel, and enhanced training for primary care physicians. Some of these advantages have already been demonstrated, while other have not yet been proved. Thus, more studies are needed to confirm the health care benefits of teledermatology.^{1,5}

A limitation of the present study, in our opinion, is that we only have histologic confirmation for patients with malignant tumors, who were a minority. However, a histologic study was performed when there was diagnostic doubt, whether the condition was inflammatory or neoplastic. These studies served to clarify the doubts (Table 5). Nonetheless,

particularly in the case of common diseases not involving a tumor, such as seborrheic dermatitis or acne, the gold standard for diagnosis in clinical practice is expert opinion rather than histology. Thus, one limitation of our study and many others is the absence of a valid measurement of interobserver reliability between dermatologists in FF consultation. This is a crucial factor in measuring diagnostic reliability, since it is possible that the interobserver disagreements (between the teledermatologist and the dermatologist who saw the patient in person) are not the result of the modalities used (teleconsultation vs FF consultation) but rather interobserver variability. The only 2 studies that included such a measurement reported widely differing results.^{17,34} While many authors have drawn attention to this problem,^{1,12,17,35} surprisingly little research has been undertaken to resolve the issue.

In conclusion, our study showed SF to be reliable in a real clinical setting. The complete independence of asynchronous SF teledermatology from the constraints of time and place make it an enormously attractive option. Our findings underscore the importance of using a standardized medical history form including mandatory items, and the need to train the physicians in digital photography. Logically, the dermatologists must be aware of the potential seriousness of the disease they are assessing. Whenever the photograph is of poor quality or their confidence in the accuracy of their own diagnosis is low, dermatologists should recommend an in-person consultation or a repeat teleconsultation. High quality evidence supporting the routine use of teledermatology as an alternative to the conventional pathways for referral of patients to specialized care has started to accumulate gradually.

Ethical Responsibilities

Protection of persons and animals. The authors declare that no experiments were performed on humans or animals during the course of this study.

Confidentiality of data. The authors declare that they have followed the protocols of their hospitals concerning the publication of patient data and that all patients included in this study were appropriately informed and gave their written informed consent.

Right to privacy and informed consent. The authors declare that no private patient data are disclosed in this article.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Funding

This study received funding through the Programa de Investigación Biomédica y en Ciencias de la Salud from the Spanish Ministry of Health and Consumer Affairs Health Research Fund (Fondo de Investigación Sanitaria) under the registration number PI021695, cofinanced by FEDER funds. It also

received financial support from the Fundación para la Investigación Sanitaria en Castilla-La Mancha (FISCAM).

References

- Hersh W, Hickam D, Severance S, Dana T, Krages K, Helfand M. Telemedicine for the Medicare Population: Update. AHRQ No. 06-E007 2006.
- Global Markets for Telemedicine Technologies. BCC Research; March 2012 [cited 2013 Jan 20]. Available from: <http://www.bccresearch.com/report/telemedicine-technologies-global-markets-hlc014e.html>
- Heffner VA, Lyon VB, Brousseau DC, Holland KE, Yen K. Store-and-forward teledermatology versus in-person visits: A comparison in pediatric teledermatology clinic. *J Am Acad Dermatol.* 2009;60:956–61.
- Romero G, Sanchez P, Garcia M, Cortina P, Vera E, Garrido JA. Randomized controlled trial comparing store-and-forward teledermatology alone and in combination with web-camera videoconferencing. *Clin Exp Dermatol.* 2010;35:311–7.
- Warshaw EM, Hillman YJ, Greer NL, Hagel EM. Teledermatology for diagnosis and management of skin conditions: A systematic review. *J Am Acad Dermatol.* 2011;64:759–72.
- Wootton R, Bloomer SE, Corbett R, Eedy DJ, Hicks N, Lotery HE, et al. Multicentre randomised control trial comparing real time teledermatology with conventional outpatient dermatological care: Societal cost-benefit analysis. *BMJ.* 2000;320:1252–6.
- Baba M, Seckin D, Kapdagli S. A comparison of teledermatology using store-and-forward methodology alone, and in combination with Web camera videoconferencing. *J Telemed Telecare.* 2005;11:354–60.
- Edison KE, Ward DS, Dyer JA. Diagnosis, diagnostic confidence, and management concordance in live-interactive and store-and-forward teledermatology compared to in-person examination. *Telemed J E Health.* 2008;14:889–95.
- Finch TL, Mair FS, May CR. Teledermatology in the UK: Lessons in service innovation. *Br J Dermatol.* 2007;156:521–7.
- Eminovic N, de Keizer NF, Bindels PJ, Hasman A. Maturity of teledermatology evaluation research: A systematic literature review. *Br J Dermatol.* 2007;156:412–9.
- Romero G, Cortina P, Vera E. Telemedicine and teledermatology (II): Current state of research on dermatology teleconsultations. *Actas Dermosifiliogr.* 2008;99:586–97.
- Du Moulin MF, Bullens-Goessens YI, Henquet CJ, Brunenberg DE, de Bruyn-Geraerds DP, Winkens RA, et al. The reliability of diagnosis using store-and-forward teledermatology. *J Telemed Telecare.* 2003;9:249–52.
- Warshaw E, Greer NL, Hillman YJ. Teledermatology for diagnosis and management of skin conditions: A systematic review of the evidence. VA-ESP Project #09-009; 2010 [cited 2013 Jan 20] Available from: <http://www.hsrd.research.va.gov/publications/esp/telederm-2010.pdf>
- Krupinski EA, LeSueur B, Ellsworth L, Levine N, Hansen R, Silvis N, et al. Diagnostic accuracy and image quality using a digital camera for teledermatology. *Telemed J.* 1999;5:257–63.
- Lyon CC, Harrison PV. Digital imaging and teledermatology: Educational and diagnostic applications of a portable digital imaging system for the trainee dermatologist. *Clin Exp Dermatol.* 1997;22:163–5.
- Kvedar JC, Menn ER, Baradagunta SL. Teledermatology in a capitated delivery system using distributed information architecture: Design and development. *Telemed J.* 1999;5:357–66.
- Whited JD, Hall RP, Simel DL, Foy ME, Stechuchak KM, Drugge RJ, et al. Reliability and accuracy of dermatologists' clinic-based and digital image consultations. *J Am Acad Dermatol.* 1999;41:693–702.
- High WA, Houston MS, Calobrisi SD. Assessment of the accuracy of low-cost store-and-forward teledermatology consultation. *J Am Acad Dermatol.* 2000;42:776–83.
- Pak HS, Harden D, Cruess Welch ML, Poropatich RD. Teledermatology: An intraobserver diagnostic correlation study, part II. *Cutis.* 2003;71:476–80.
- Taylor P. An assessment of the potential effect of a teledermatology system. *J Telemed Telecare.* 2000;6 Suppl 1:574–6.
- Zelickson BD, Homan L. Teledermatology in the nursing home. *Arch Dermatol.* 1997;133:171–4.
- Bowns IR, Collins K, Walters SJ, McDonagh AJ. Telemedicine in dermatology: A randomised controlled trial. *Health Technol Assess.* 2006;10:1–39.
- Mahendran R, Goodfield MJ, Sheehan-Dare RA. An evaluation of the role of a store-and-forward teledermatology system in skin cancer diagnosis and management. *Clin Exp Dermatol.* 2005;30:209–14.
- Moreno-Ramirez D, Ferrandiz L, Nieto-Garcia A, Carrasco R, Moreno-Alvarez P, Galdeano R, et al. Store-and-forward teledermatology in skin cancer triage: Experience and evaluation of 2,009 teleconsultations. *Arch Dermatol.* 2007;143:479–84.
- Lim AC, Egerton IB, See A, Shumack SP. Accuracy and reliability of store-and-forward teledermatology: preliminary results from the St. George Teledermatology Project. *Australas J Dermatol.* 2001;42:247–51.
- Romero G, Garrido JA, Garcia-Arpa M. Telemedicine and teledermatology (I): Concepts and applications. *Actas Dermosifiliogr.* 2008;99:506–22.
- Wootton R, Bahaadinbeigy K, Hailey D. Estimating travel reduction associated with the use of telemedicine by patients and healthcare professionals: Proposal for quantitative synthesis in a systematic review. *BMC Health Serv Res.* 2011;11:185–94.
- Ferrandiz L, Moreno-Ramirez D, Ruiz-de-Casas A, Nieto-Garcia A, Moreno-Alvarez P, Galdeano R, et al. An economic analysis of presurgical teledermatology in patients with nonmelanoma skin cancer. *Actas Dermosifiliogr.* 2008;99:795–802.
- Van der Heijden JP, de Keizer NF, Bos JD, Spuls PI, Witkamp L. Teledermatology applied following patient selection by general practitioners in daily practice improves efficiency and quality of care at lower cost. *Br J Dermatol.* 2011;165:1058–65.
- Pak HS, Datta SK, Triplett CA. Cost minimization analysis of a store-and-forward teledermatology consult system. *Telemed J E Health.* 2009;15:160–5.
- Pak H, Triplett CA, Lindquist JH. Store-and-forward teledermatology results in similar clinical outcomes to conventional clinic-based care. *J Telemed Telecare.* 2007;13:26–30.
- Ferrandiz L, Ruiz-de-Casas A, Martin-Gutierrez FJ, Peral-Rubio F, Mendez-Abad C, Rios-Martin JJ, et al. Effect of teledermatology on the prognosis of patients with cutaneous melanoma. *Arch Dermatol.* 2012;148:1025–8.
- Warshaw EM, Lederle FA, Grill JP, Gravely AA, Bangerter AK, Fortier LA, et al. Accuracy of teledermatology for pigmented neoplasms. *J Am Acad Dermatol.* 2009;61:753–65.
- Leshner JL Jr, Davis LS, Gourdin FW. Telemedicine evaluation of cutaneous diseases: A blinded comparative study. *J Am Acad Dermatol.* 1998;38:27–31.
- Nordal EJ, Moseng D, Kvammen B, Nochen ML. A comparative study of teleconsultations versus face-to-face consultations. *J Telemed Telecare.* 2001;7:257–65.