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; 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 < 0.001), diagnostic confidence (P < 0.001), felt need for conventional consultation (P < 0.001), and the quality of the clinical record (P = 0.013).


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

$1578-2190$ - see front matter © 2013 Elsevier España, S.L. and AEDV. All rights reserved.
Introduction

A growing number of patients, physicians, and institutions are using telemedicine services, and the world telemedicine market is expected to grow from US$9.8 billion in 2010 to some US$27.3 billion in 2016. Dermatology appears to be an ideal specialty for telemedicine. Indeed, it is the clinical specialty that currently accounts for the largest number of telemedicine studies. 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. It can also be more costly than conventional face-to-face (FF) consultation. 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. Thus, medical professionals and healthcare providers increasingly consider SF to be the most cost effective and convenient system.

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, 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. 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%. 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.

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...
by experts, whether dermatologists, dermatology residents, professional photographers, or research assistants. 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, 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.

#### 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 χ² test for 2 × 2 tables. To ascertain whether the tele dermatologists 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 to 86). 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 (χ² = 2.13, P = .334), or diagnostic category (χ² = 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 FF 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).

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).

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
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.1123 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 2  Diagnostic Agreement: Intermodality (SF vs FF) and Intramodality Between the 3 Teledermatologists (SF vs SF).^a

<table>
<thead>
<tr>
<th>Diagnostic Agreement</th>
<th>Intermodality: SF vs FF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF D2-FF D1  n = 170</td>
</tr>
<tr>
<td>Complete agreement</td>
<td>72.9</td>
</tr>
<tr>
<td>Partial agreement</td>
<td>15.3</td>
</tr>
<tr>
<td>Disagreement</td>
<td>11.8</td>
</tr>
</tbody>
</table>

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

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

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

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

<table>
<thead>
<tr>
<th>Modifying Factors</th>
<th>Diagnostic Agreement, %</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete Agreement</td>
<td>Partial Agreement</td>
</tr>
<tr>
<td>Diagnostic Confidence</td>
<td>Low</td>
<td>47.2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>73.5</td>
</tr>
<tr>
<td>Need for face-to-face consultation</td>
<td>Yes</td>
<td>43.1</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>78.3</td>
</tr>
<tr>
<td>Image quality</td>
<td>Poor</td>
<td>57.6</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>74.1</td>
</tr>
<tr>
<td>Quality of medical history</td>
<td>Poor</td>
<td>46.7</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>71.3</td>
</tr>
</tbody>
</table>
Table 5  Illustrative Instances of Discrepancies Between the 4 Assessments: 3 Teleconsultations (D1, D2, D3) and 1 Face-to-Face Consultation (D1).

<table>
<thead>
<tr>
<th>Case No</th>
<th>FF D1</th>
<th>SF-D1</th>
<th>SF-D2</th>
<th>SF-D3</th>
<th>AP</th>
<th>FF-R</th>
<th>DC</th>
<th>PQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>Pilomatrixoma</td>
<td>Dermatofibroma vs epidermal cyst</td>
<td>Epidermal Cyst</td>
<td>Dermatofibroma vs angioleiomyoma</td>
<td>Juvenile xanthogranuloma</td>
<td>Yes</td>
<td>Low</td>
<td>Poor</td>
</tr>
<tr>
<td>Case 2</td>
<td>Pityriasis rosea</td>
<td>Seborrheic dermatitis</td>
<td>Acute eczema vs seborrheic dermatitis</td>
<td>Seborrheic dermatitis</td>
<td>NA</td>
<td>Yes</td>
<td>Low</td>
<td>Poor</td>
</tr>
<tr>
<td>Case 3</td>
<td>Nevis depigmentosus</td>
<td>Lichen sclerosus et atrophicus vs vitiligo</td>
<td>Vitiligo vs lichen sclerosus et atrophicus</td>
<td>Vitiligo</td>
<td>NA</td>
<td>Yes</td>
<td>Low</td>
<td>Poor</td>
</tr>
<tr>
<td>Case 4</td>
<td>Granuloma annulare vs common wart</td>
<td>Carcinoma of the tongue</td>
<td>Carcinoma of the tongue</td>
<td>Eosinophilic ulcer</td>
<td>Carcinoma of the tongue</td>
<td>Yes</td>
<td>Low</td>
<td>Poor</td>
</tr>
<tr>
<td>Case 5</td>
<td>Alopecia in discoid lupus</td>
<td>Alopecia areata</td>
<td>Scarring alopecia</td>
<td>Alepecia</td>
<td>Discoid lupus</td>
<td>Yes</td>
<td>Low</td>
<td>Poor</td>
</tr>
<tr>
<td>Case 7</td>
<td>Drug-induced LE</td>
<td>Drug-induced LE</td>
<td>Subacute LE vs drug induced LE</td>
<td>Photodermatitis</td>
<td>Lupus erythematous</td>
<td>Yes</td>
<td>High</td>
<td>Good</td>
</tr>
</tbody>
</table>

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.

Most earlier studies have assessed the reliability of SF under ideal “laboratory” conditions, unlike those of routine workflow in a normal clinical setting. In the study by Krupinski et al., the photographs were taken by a medical student under the direction of a dermatologist, and in the study by Lyon et al., the photographs were taken by a resident in dermatology. In one study, the photographs were taken by a professional photographer, and in others by a research assistant. Finally, in a number of other studies, photographs were taken by a nurse trained in clinical photography. 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. This procedure made it easier to carry out the study but may have introduced bias.

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), morpha (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)\(^a\)

<table>
<thead>
<tr>
<th>Treatment Agreement</th>
<th>Intermodality: SF vs FF (Validity)</th>
<th>Intramodality SF vs SF (Reliability)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF D2-FF D1(^b) n = 170</td>
<td>SF D3-FF D1 n = 170</td>
</tr>
<tr>
<td>Complete agreement</td>
<td>57.1</td>
<td>65.9</td>
</tr>
<tr>
<td>Partial agreement</td>
<td>20.0</td>
<td>17.6</td>
</tr>
<tr>
<td>Disagreement</td>
<td>22.9</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>Intramodality SF vs SF (Reliability)</td>
<td>SF D1-SF D2 n = 170</td>
</tr>
<tr>
<td>Complete agreement</td>
<td>44.1</td>
<td>52.4</td>
</tr>
<tr>
<td>Partial agreement</td>
<td>17.1</td>
<td>25.9</td>
</tr>
<tr>
<td>Disagreement</td>
<td>38.8</td>
<td>21.8</td>
</tr>
</tbody>
</table>

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.\(^{34}\) 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.\(^{7,14,16}\)

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,\(^{7,14,16}\) the present study was the first to analyze the correlation between diagnostic agreement and the decision to recommend an in-person consultation.

Our statistical analysis showed with a high degree of statistical significance that the teledermatologist can predict and therefore avoid errors in diagnosis (\(P < 0.001\) 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 nevi; 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. 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.

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. We believe that our study and other recent studies provide evidence to support the usefulness of teledermatology in routine practice. One randomized clinical trial showed that SF can be a cost-saving strategy for providing dermatological care, producing clinical outcomes similar to those obtained with conventional consultations. A recent study demonstrated the feasibility of SF on a large scale: 1820 general practitioners and 166 dermatologists performed more than 37000 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. 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, 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. 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 others have not yet been proved. Thus, more studies are needed to confirm the health care benefits of teledermatology.

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. While many authors have drawn attention to this problem, 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 makes 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


