Telemedicine and Teledermatology (II): Current State of Research on Dermatology Teleconsultations

G. Romero, P. Cortina, and E. Vera
Unidad de Dermatología, Hospital General de Ciudad Real, Ciudad Real, Spain

Abstract. The rapid increase in the use of telemedicine makes a critical appraisal of these services essential. This article reviews the current evidence from studies of teledermatology with sufficiently rigorous methodology to allow reasonable conclusions to be drawn. The review is organized according to areas of research in teledermatology: reliability of diagnosis and management, accuracy, outcomes, costs, and satisfaction. The conclusion of this review is that teledermatology has been shown to be feasible and reliable but that it has a long way to go before maturing as a medical technology in daily practice. The largest body of evidence relates to the reliability of diagnosis and management, but although numerous studies have been performed, many have marked methodological shortcomings. Accuracy studies have been conducted in skin oncology with appropriate methodology that show an accuracy comparable to that of face-to-face consultations. Studies of health outcomes should be designed such that a faithful assessment of the final outcomes, costs, and satisfaction with teledermatology systems can be carried out.

Key words: telemedicine, teleconsultation, teledermatology, store-and-forward systems, real-time systems, reproducibility of results, validation studies, outcomes assessment, costs analysis, personal satisfaction.

As we reported in the first part of this review, teledermatology has many applications that undoubtedly have the potential to revolutionize dermatological care. However, the question remains, has the efficacy, effectiveness, and efficiency of teleconsultation been adequately demonstrated? Since the mid-1990s, a number of authors have written about the real possibilities of the practice of telemedicine as well as the risks and difficulties involved. Although teledermatology is a young discipline, a great deal of research has been undertaken in this area in recent years, making it the most studied clinical discipline in telemedicine and the one with the greatest potential influence on health care policy decisions. However, the value of the research undertaken to date is the subject of some debate, and recently the quality of the...
are comparable to those produced by the clinic-based process. The aim of this review is to shed some light on this debate, as it is clearly of current interest and importance.

There are 5 important areas in the field of teledermatology research and each one must be discussed separately:

1. Analysis of the reliability of teleconsultation in terms of the reproducibility of results. Will different dermatologists evaluating the same data and clinical images reach the same conclusions regarding the diagnosis and management of the patient’s condition?
2. Analysis of the validity or diagnostic accuracy of the teleconsultation. Are the diagnostic conclusions reached using teledermatology as accurate as those reached in the course of a clinic-based face-to-face consultation?
3. Analysis of results. Are the clinical outcomes achieved when a case is managed via teleconsultation similar to those obtained in a conventional clinic-based consult process?
4. Cost analysis. Is teleconsultation more cost-effective than conventional in-person consultation?
5. Satisfaction assessments. Are the participants (referring clinicians, dermatologists, and patients) as satisfied with the teleconsultation process as they are with the conventional clinic-based process?

Reliability Studies in Teledermatology

Studies of diagnostic agreement account for the largest body of evidence and greatest research output in the field of teledermatology. This is not surprising since the assessment of reliability is the key component in the evaluation of any technology, and medical diagnostic procedures are no exception. The fundamental question that must be answered is whether the diagnostic decisions and management plans produced by the teleconsultation process are comparable to those produced by the clinic-based process.

Reliability and Validity

In clinical medicine, reliability usually refers to the repeatability of a measurement. Consequently, we measure the degree of agreement between different assessments of the same case. In this context, the term reliability is used as a synonym for repeatability, reproducibility, or agreement. The term validity refers to whether the procedure is really measuring the phenomenon we want to measure, that is, whether the measurement or outcome of the assessment is accurate. Since we cannot assess the validity of an instrument without first establishing its reliability, before investigating whether the instrument measures what we want to measure, we must ascertain whether it reliably measures something, that is, whether the measurement can be reproduced. An instrument that does not yield reproducible results is unreliable and consequently any assessment of the accuracy of the measurement is pointless.

In teledermatology, the experimental designs used to assess reliability are based on the study of repeated measurements. A dermatologist makes a diagnostic or management decision concerning a clinical problem, and then the same clinician (in intraobserver agreement studies) or a different clinician (in interobserver studies) repeats the assessment, and the agreement between the 2 decisions is analyzed. If concordance is high (a simple percentage of agreement of more than 80% or K>0.6), the instrument is deemed reliable.

Thus, to properly assess the validity of teleconsultation, the results obtained using this method must be compared with those obtained using a second instrument (a gold standard). In dermatology, however, there may not be any well established gold standard. Histology is used as the gold standard in neoplastic disease, microbiology in infectious disease, skin testing in contact dermatitis, and the clinical-pathological correlation in inflammatory disease.

Results of Diagnostic Reliability Studies

In teledermatology, most authors use clinic-based face-to-face consultation as the gold standard, since they consider that patient assessment using the conventional in-person process will always be equal to or better than assessment via teleconsultation. The most common experimental design in teledermatology is the comparison of the diagnoses obtained via teleconsultation with those obtained after clinic-based consultation. These studies assess both reliability and validity. This design is, however, merely an approximation, and these studies would be more valuable if the results could be compared to a real gold standard, which could be the correlation between the clinical findings resulting from a face-to-face visit and the results of the pertinent additional investigations in each case (histological studies, cultures, patch tests, etc).

Diagnostic agreement is generally categorized as being either complete or partial. Agreement is defined as complete when both examiners establish the same diagnosis. Partial agreement refers to a situation in which at least 1 of the 2 specialists develops a differential diagnosis that includes at least 1 of the diagnoses specified by the other specialist. Many studies cite the value for aggregate agreement, which is the sum of complete and partial agreement.

The results of almost all of the studies that used the most common experimental design (comparison of teleconsultation with in-person consultation) report an acceptable level of diagnostic reliability for teleconsultation (>0.6). The results do, however, vary considerably. In interobserver studies of real-time video conferencing systems, complete diagnostic agreement ranged from 0.54 to 0.80 while aggregate agreement ranged from 0.79 to 0.99. In studies of store-and-forward teledermatology (asynchronous processes based
A group of researchers in Oregon published systematic reviews in 2001 and 2006 analyzing the level of evidence available on the clinical applications of telemedicine, with the exclusion of teleradiology and telepathology (Table 3).

They classified the results of the studies according to the following system based on the findings regarding direction of effect: A, strong positive effect; B, weak positive effect; C, conflicting evidence; and D, negative effect (evidence that the technology is inferior or ineffective).

In the first review, 18 of the 55 studies of diagnosis and management in the various specialties dealt with teledermatology. Half of the 22 studies on store-and-forward systems referred to teledermatology, and the quality scores of these 11 studies were as follows: 2 were I-B,

6 were II-B, 3,35,36,40,42 2 were II-C, 34,43 and 1 was III-C. 44 Of the 33 studies of real-time systems, 7 dealt with teledermatology (2 classified as II-C, and 5 as III-C,23,25,27,45,46). In the 2006 update, 14 of the 52 studies of diagnosis and management concerned teledermatology. Of these, 13 were on store-and-forward systems (2 classified as I-B, 47,48 9 as II-B, 30,31,49-55 and 2 as II-C,57,58) and only 1 dealt with real-time teledermatology. Most of these authors reported a weak positive effect (B) in favor of teledermatology, and none of them reported a negative effect (D). Nevertheless, none of the authors reported a strong positive effect (A). The 2 studies with the highest quality classification dealt with the diagnosis of pigmented lesions and included a measurement of validity (histological gold standard) for both the teleconsultation and clinic-based groups. 47,48

Dermatology, which is in the forefront of clinical telemedicine research in the area of diagnosis and management, accounted for 32 of the 107 studies reviewed (29%). Interest in real-time teledermatology based on interactive video conferencing has declined among researchers. In the period covered by the first review, 7 of the 22 studies dealt with real-time teleconsultation as compared to only 1 of the 13 studies reviewed in the following 5-year period (2001-2006). Interobserver concordance was the aspect of teledermatology most often

### Table 1. Interobserver Reliability in Real-Time Teledermatology I

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Patients (Consultations)</th>
<th>Valid Cases</th>
<th>CA (N)</th>
<th>AA (N)</th>
<th>D (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phillips, 1997</td>
<td>60 (79)</td>
<td>79</td>
<td>0.77 (61)</td>
<td>–</td>
<td>0.23 (18)</td>
</tr>
<tr>
<td>Gilmour, 1998</td>
<td>126 (155)</td>
<td>76a</td>
<td>0.54 (41)</td>
<td>0.80 (61)</td>
<td>0.20 (15)</td>
</tr>
<tr>
<td>Lesher, 1998</td>
<td>60 (68)</td>
<td>68</td>
<td>0.78 (53)</td>
<td>0.99 (67)</td>
<td>0.06 (1)</td>
</tr>
<tr>
<td>Lowitt, 1998</td>
<td>112 (130)</td>
<td>130</td>
<td>0.80 (104)</td>
<td>–</td>
<td>0.20 (26)</td>
</tr>
<tr>
<td>Loane, 1998</td>
<td>351 (427)</td>
<td>155a</td>
<td>0.60 (93)</td>
<td>0.79 (122)</td>
<td>0.21 (33)</td>
</tr>
<tr>
<td>Nordal, 2001</td>
<td>121 (112)</td>
<td>112</td>
<td>0.72 (81)</td>
<td>0.86 (97)</td>
<td>0.13 (15)</td>
</tr>
<tr>
<td>Aggregate total</td>
<td>830 (971)</td>
<td>620a</td>
<td>0.698 (433)</td>
<td>0.844 (347)</td>
<td>0.174 (108)</td>
</tr>
</tbody>
</table>

*aCases measuring intraobserver agreement were eliminated. Abbreviations: AA, aggregate agreement; CA, complete agreement; D, disagreement; N, number in sample.

### Table 2. Studies of Interobserver Diagnostic Reliability in Store-And-Forward Teledermatology

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Patients (Consultations)</th>
<th>Cases</th>
<th>CA</th>
<th>AA</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kvedar, 1997</td>
<td>116</td>
<td>242</td>
<td>0.62</td>
<td>0.68</td>
<td>0.31</td>
</tr>
<tr>
<td>Lyon, 1997</td>
<td>100</td>
<td>100</td>
<td>0.90</td>
<td>–</td>
<td>0.10</td>
</tr>
<tr>
<td>Zelickson, 1997</td>
<td>30</td>
<td>60</td>
<td>0.88</td>
<td>–</td>
<td>0.12</td>
</tr>
<tr>
<td>Whited, 1999</td>
<td>129</td>
<td>1008a</td>
<td>0.47</td>
<td>0.86</td>
<td>0.14</td>
</tr>
<tr>
<td>High, 2000</td>
<td>92</td>
<td>275</td>
<td>0.70</td>
<td>0.85</td>
<td>0.15</td>
</tr>
<tr>
<td>Lim, 2001</td>
<td>49</td>
<td>212</td>
<td>0.79</td>
<td>0.86</td>
<td>0.14</td>
</tr>
<tr>
<td>Taylor, 2001</td>
<td>194</td>
<td>101b</td>
<td>0.50</td>
<td>0.60</td>
<td>0.39</td>
</tr>
<tr>
<td>Du Moulin, 2003</td>
<td>117</td>
<td>117</td>
<td>0.54</td>
<td>0.63</td>
<td>0.37</td>
</tr>
<tr>
<td>Baba, 2005</td>
<td>228</td>
<td>242</td>
<td>0.75</td>
<td>–</td>
<td>0.25</td>
</tr>
<tr>
<td>Aggregate total</td>
<td>849</td>
<td>2444</td>
<td>0.60</td>
<td>0.80</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*The participation of 6 observers increased the power of this study. Abbreviations: AA, aggregate agreement; CA, complete agreement; D, disagreement.

### Table 3. Classification of Studies by Methodological Quality in the Review by Hersh et al

<table>
<thead>
<tr>
<th>Class</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Case series of consecutive patients from relevant population of individuals who would use teledermatology; using an objective gold standard with blinded interpretation of results; with interobserver analysis;</td>
</tr>
<tr>
<td>II</td>
<td>Case series of patients from relevant population of individuals who would use teledermatology; using an objective gold standard</td>
</tr>
<tr>
<td>III</td>
<td>Case series not from relevant population or not using appropriate methodology for diagnostic test evaluation</td>
</tr>
</tbody>
</table>

on the use of static digital images), complete agreement ranged from 0.47 to 0.90 and aggregate agreement from 0.59 to 0.86 (Tables 1 and 2).
evaluated, and the findings varied widely (0.41–0.87 for complete agreement and 0.51–0.96 for aggregate agreement). With the exception of 2 studies in the first review,\textsuperscript{26,28} the striking omission was that none of the studies reviewed measured baseline agreement for either diagnosis or management in face-to-face consultation. The authors of these reviews are of the opinion that the only way to clarify why diagnostic concordance varies is to design high-quality studies that include baseline assessment of interobserver agreement in face-to-face consultation.

During the last 2 years since the 2006 review, interest has centered on the management of tumors\textsuperscript{57} (especially pigmented lesions\textsuperscript{58}) and teledermoscopy.\textsuperscript{59,60} Teledermoscopy will be discussed in more detail in the section on validity.

### Disease Management Reliability Studies

Very few authors have assessed management in terms of the treatment prescribed and the additional tests recommended. They found the reliability of management decisions taken on the basis of a teleconsultation to be high, as in the case of diagnosis. In the only study that used in-person consultation as a baseline,\textsuperscript{26} only 1 of the 3 teledermatologists had a lower reliability index for management decisions than that obtained in face-to-face consultation. Apart from this exception, no significant differences were found between clinic-based examination and teleconsultation in the reliability of recommendations for medical or surgical treatment. Diagnosis was generally found to be more reliable than management, and the authors attributed this difference to the fact that management reflects individual preferences while diagnosis is more strictly guided by criteria. The authors of a study that evaluated a store-and-forward system found correct management in 90% of cases when the examiner was given both an image and a clinical history and in 87% of cases when only an image was supplied.\textsuperscript{38}

In the only 2 studies that assessed the reliability of management in real-time teleconsultation, it was found to be correct in over 70% of cases.\textsuperscript{25,27} In general, the same number and type of diagnostic tests were recommended by clinic-based dermatologists and teledermatologists who used a video link.\textsuperscript{28} Diagnostic agreement was found to correlate with management agreement ($P<.0001$).\textsuperscript{29} In another study, complete agreement on the tests recommended was reported in 88% of cases and on treatment in 84% of cases, but the value of these results is lower because of the intraobserver design of the study.\textsuperscript{61}

Another area that has been studied is the reliability of the decision to take a biopsy. The results of this research have been variable, with reliability ranging from 0.45\textsuperscript{40} to 1.00\textsuperscript{55} in 2 studies that analyzed asynchronous systems. In another study of a store-and-forward system, Pak et al\textsuperscript{42} found intraobserver reliability for establishing an indication for biopsy to be 0.76 ($\kappa=0.47$), a finding very similar to that of Phillips et al\textsuperscript{35} in real-time teledermatology (0.86; $\kappa=0.47$).

The authors of a recent study that analyzed presurgical assessment of skin tumors found a very high level of reliability ($\kappa=0.75$; 95% confidence interval, 0.71–0.79) for the indication of the need for surgery when patients were assessed remotely on the basis of a static digital image.\textsuperscript{62}

### The Influence of Image Quality on Diagnostic Reliability

The quality of the images used has a decisive influence on diagnostic concordance. This relationship has been analyzed, and a number of studies report the same effect.\textsuperscript{30,33,36–38,63} Only 1 study reported that image quality had no effect on levels of diagnostic agreement,\textsuperscript{39} but in that case the digital photographs were taken on site by previously trained medical students working directly under the direction of a dermatologist and following a strict protocol. Consequently quality images were obtained, and there was a high level of complete agreement between the different examiners (>80%). In other studies, however, when photographs were not taken in this ideal setting but rather under normal working conditions by different physicians in the primary care center, the quality of the images was not as good and the rate of agreement was substantially lower.\textsuperscript{37,64}

Today, file size and the resolution of the digital image have ceased to be important issues, and any camera in the low-to-middle price range can be used to obtain the necessary resolution for good visualization on a standard computer screen (1024 $\times$ 768 pixels). In fact, 1 study found no differences between diagnoses based on high-quality color slides and those based on low-quality digital images; although in the same study reliability was lower when video images were used as the basis for examination.\textsuperscript{65} Another study reported no differences in diagnostic accuracy between high-quality slides and digital images compressed at ratios of 15:1, 30:1, and even 40:1.\textsuperscript{66} A third study found no differences in diagnosis between in-person consultations and the use of slides or digital photographs, but marked differences were found when the examiner was a primary care physician rather than a dermatologist.\textsuperscript{67}

Some authors have attributed greater importance to the framing of the shot than to the quality of the photograph, and specify that the image should include both unaffected and affected areas of the skin to facilitate clinical assessment.\textsuperscript{68,69} Focussing problems are common in inexperienced photographers, and sharp focus is essential to the diagnosis. Proper focus can be assured using the macro focus lens feature currently generally available on mid-range cameras. The telephoto zoom of the macro lens should not be abused, and it is important to place an object, for example a ruler, close to the skin to facilitate focusing when the skin lesion lacks contrast. Focus should always be checked on the display by enlarging the image immediately after the shot has been.
Influence of the Quality of the Clinical History

It is possible that the quality of the clinical history used does not play a key role in the diagnosis of dermatological disease. Zelickson and Homan\(^\text{36}\) reported diagnostic agreement in 67% of cases on the basis of a clinical history alone, in 85% when only a photograph was used, and in 88% when diagnosis was based on both a clinical history and a photograph. The authors of a study of hospitalized patients reported that agreement increased from 52% when only a clinical photograph was used to 65% when this was complemented by a clinical history.\(^\text{29}\) In another study, 8 dermatologists assessed 50 patients on the basis of static digital images alone.\(^\text{41}\) When a clinical history was made available, the teleconsultants changed their diagnosis in only 11% of the cases. Baba et al.\(^\text{18}\) reported that the rate of complete agreement varied between 0.73 and 0.89 depending on the quality of the clinical history. In another study, agreement was significantly lower (\(P=0.0002\)) when the quality of the clinical history was poor.\(^\text{30}\) Finally, other authors have emphasized the need for a standardized clinical history to ensure a minimum level of quality.\(^\text{26,28,68,70}\)

Diagnostic Confidence Level

As with the influence of the quality of clinical photographs, the level of diagnostic confidence (certainty) has often been the subject of investigation in the literature. The findings are consistent because all the studies report a correlation between diagnostic certainty and agreement in both asynchronous\(^\text{21,26,33}\) and real-time\(^\text{22,25,27,46}\) teledermatology. On analysis, the level of diagnostic certainty was found to be better for in-person consultation than for teleconsultation,\(^\text{22,25,33}\) although the difference reported in a Norwegian study was not statistically significant.\(^\text{22}\)

Some authors used the percentage of cases in which a differential diagnosis was established as a measure of diagnostic confidence on the basis that this percentage reflects the difficulty of assessing the teleconsultation data.\(^\text{28,25}\) In their opinion, the fact that the teleconsultants developed a differential diagnosis in more cases than the clinic-based examiners was an indication of less diagnostic certainty in both asynchronous and real-time teledermatology. Whited et al.\(^\text{26}\) and Du Moulin et al.\(^\text{17}\) found only slight differences between teledermatology and in-person consultation in the percentage of cases with a differential diagnosis and reported greater diagnostic confidence when the evaluation was performed in person. While the development of a differential diagnosis may be an indirect measurement of the level of diagnostic confidence it is, nonetheless, a tool that all clinicians use to reach a final diagnosis. Consequently, it is a measure more of the difficulty of achieving a diagnosis in the process under study, and the final diagnosis may only be pending the results of a test that will resolve the dilemma. In other words, the clinician who develops a differential diagnosis may be certain that he or she is following correct diagnostic procedure. Thus, the consultant’s own opinion is a better measure of diagnostic confidence.

In any case, the important point in this respect is that a strong correlation has been found between a high degree of diagnostic agreement and a high level of diagnostic certainty and conversely between lack of agreement and low certainty. The implication of this is that dermatologists are aware of their confidence or lack thereof, and can identify cases susceptible to misdiagnosis and take the decision that the assessment of a particular case may require an in-person consultation, further photographs, or a more complete clinical history in order to eliminate the possibility of error.\(^\text{21}\) Ideally, in the protocol applied to the use of teleconsultation for triage, 1 of the criteria for nonreferral should be a high level of diagnostic confidence, thus ensuring a sensitivity close to 1 for teleconsultation, particularly when this process is used to screen for skin cancer.\(^\text{71}\)

Diagnostic Reliability by Disease

Whether diagnostic agreement in teledermatology is greater in some diseases than others is a much debated question. In some studies, no differences were found between disease categories when agreement was analyzed according to the reason for the consultation.\(^\text{23,33}\) Most authors report a lower rate of reliability in cases of eruptions than in tumors.\(^\text{34,35,42,52}\) Given the ease of diagnosing acne and common warts, these conditions were excluded from the analysis in 1 study.\(^\text{31}\) Other authors have reported difficulties in the diagnosis of benign tumors via teledermatology,\(^\text{36,46}\) although 1 study reported nearly perfect sensitivity and specificity in such cases.\(^\text{22}\) In pigmented lesions, examiners using store-and-forward teledermatology achieved very high diagnostic agreement.\(^\text{35,68}\) The results of the 2 most recent studies of the use of store-and-forward systems for assessing patients with suspected skin cancer are inconsistent.\(^\text{57,64}\) While the authors of a large study of 2009 patients reported interobserver concordance of \(k=0.83\), the examiners in a study involving 163 patients in the United Kingdom did not reach a simple concordance of 0.50. However, in the British study, 20% of the photographs...
supplied were deemed to be of inadequate quality as compared to only 6% of those used in the large series.

Methodological Problems in the Diagnostic Reliability Studies Analyzed

Important shortcomings prevent us from drawing firm conclusions about the reliability of teledermatological consultation. Most of these defects concern methodological and experimental design problems that make it difficult to extrapolate the results obtained to the target population with a minimum of methodological rigor. The main problems are as follows:

Almost Complete Lack of Any Randomized Trials

Patients were generally included consecutively or the criteria used were not described. Only 3 randomized reliability studies were found,28,31,52 Moreover, in 2 of these28,52 the authors did not explain how the randomization was carried out, and in the third, the sample population was very small (72 patients, 36 randomized to store-and-forward teledermatology and 36 to clinic-based consultation).31 Some cost analyses have been based on groups formed using a proper randomization method, but these studies did not analyze reliability,12-25 except 1 study that compared reliability in a group of patients who were assessed twice (once using a store-and-forward system and once via real-time teledermatology),26 but this study did not compare these modalities with the gold standard of in-person consultation. A recent study of store-and-forward teledermatology used a randomized prospective design, but recruitment problems and losses to follow-up during the trial reduced the value of the results obtained.27

No Control Group

One study of real-time teledermatology included a small control group of 29 patients who were diagnosed in a face-to-face consultation.22 However, the authors of that study failed to provide the necessary demographic and diagnostic control-group data that would provide a basis for comparison with the group of 112 patients treated via teleconsultation. Nor did they measure diagnostic reliability in the control group.

Only 2 studies used the ideal design, that is, one including a control group of patients assessed in person and measurement of the baseline interobserver reliability of the teleconsultants in face-to-face consultation. In only 1 study of store-and-forward teledermatology, a group of 129 patients (with 168 dermatological conditions) were assessed by 2 dermatologists in person (making it possible to measure interobserver reliability) and by 3 other dermatologists in teleconsultation.26 This is really the only study that offers the possibility of a valid comparison between the reliability of teleconsulting and that of conventional face-to-face consultation. A study of real-time teledermatology included a control group of 36 patients with 47 dermatological conditions. These patients were assessed by the teledermatologists in face-to-face consultations.29 However, the value of this study is limited by the small sample size.

Moreover, the results of the only 2 studies that included controls who were assessed in person are inconsistent, with complete agreement after analysis of 165 patients in face-to-face consultation of 0.54 in 1 study and 0.94 in the other (Table 4). These results contrast with the more than 30 studies that analyzed the reliability of the different modalities of teleconsultation in over 2000 dermatological patients. However, the lack of a valid measurement of interobserver reliability in face-to-face consultation is a crucial factor, since it makes it impossible to compare the reliability observed in teleconsultation with a control baseline.

It is possible that disagreement between 2 observers (a teledermatologist and a clinic-based consultant) may not be a result of the technology used in each case but rather simply due to interobserver variability, the value of which is unknown in these studies. While many authors have called attention to this problem,12,22,26,32,37,78 surprisingly little research activity has been undertaken to resolve the question.

Some authors have tried to get around this problem by using an intraobserver study design.52 This design eliminates interobserver variability since the same dermatologist establishes the diagnosis twice (once on the basis of a teleconsultation and once after an in-person consultation). However, it introduces a carry-over bias since the dermatologist is aware of the previously established diagnosis and this gives rise to a falsely high rate of reliability. On analysis of the intraobserver studies on asynchronous25,38,42,52 and real-time dermatology,25,38,79 we observed that the indices of reliability obtained were significantly higher than those reported in interobserver studies. Authors who analyzed both interobserver and intraobserver reliability in the same sample population have found significantly higher reliability in the intraobserver analysis than in the interobserver modality.27,38,63,79 Although it cannot be eliminated, this bias can be minimized by introducing a washout period between the 2 consultations so that the dermatologists may not recall their original assessment of the case. Since we have no design that would quantify the bias introduced by the intraobserver model, the only way of clarifying the situation is to measure the interobserver variability.

### Table 4. Interobserver Reliability in Conventional Face-to-Face Consultation

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Patients</th>
<th>Dermatologists</th>
<th>Cases</th>
<th>CA</th>
<th>AA</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesher, 1998</td>
<td>36</td>
<td>2</td>
<td>47</td>
<td>0.94</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Whitel, 1999</td>
<td>129</td>
<td>2</td>
<td>168</td>
<td>0.54</td>
<td>0.92</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Abbreviations: AA, aggregate agreement; CA, complete agreement; D, disagreement.
in face-to-face consultation within the group of participating examiners, thereby providing a valid baseline for the subsequent comparison with the results obtained after teleconsultation.

**Predominance of Studies of Efficacy Rather Than Effectiveness in Store-and-Forward Teledermatology**

Most of the studies have studied efficacy in ideal “laboratory” conditions rather than effectiveness in a situation comparable to that of the routine workflow in a normal clinic. The bias introduced particularly affects studies of store-and-forward processes. In routine practice, the primary care physician or trained health care professional sees the patient, records the clinical information, and takes the photographs that are sent over an intranet or the Internet and assessed remotely by the teledermatologist. In the study published by Kvedar et al, a professional photographer took the photographs, which were then assessed on-site by a dermatologist rather than being sent to a remote location. In 2 other studies, the photographs were taken by a dermatology intern, and in a third by a medical student under the direction of a dermatologist. In both cases, the teleconsultant’s diagnosis was facilitated by the fact that a dermatologist decided which areas should be photographed. In the studies by High et al and Whited et al, a research assistant took the photographs and recorded the clinical information. In a series of other studies, a nurse trained in clinical photography performed these tasks. In most of the studies, the photographs were taken and stored on-site. When primary care physicians have taken the photographs and sent the files to the remote location, the results are clearly not as good as the average quality achieved in the studies cited above. The only times high levels of reliability in diagnosis and management were reported when primary care physicians were responsible for presenting the store-and-forward teleconsultation were in the studies in skin cancer triage undertaken by Moreno-Ramírez et al. This raises the question of whether the real level of reliability for asynchronous teledermatology in general consulting studied by way of an adequate experimental design measuring the real-life effectiveness of teleconsulting under normal working conditions might not be lower than that indicated by the results in the literature. This problem is less important in real-time teledermatology because the dermatologist directs the teleconsultation and the role of the primary care physician is secondary.

**Validity Studies**

The only studies that compared telediagnosis with a, generally histological, gold standard investigated store-and-forward teledermatology. In these studies on general dermatological disease, researchers have analyzed validity in subgroups of patients for whom biopsy results were available. Once again, the validity of store-and-forward teledermatology was variable, ranging from 0.63 to 1.00. In the 2 studies that provide validity data for in-person consultation, the validity of face-to-face assessment was reported to be higher than that of store-and-forward teledermatology, with indexes of 0.89 vs 0.76 in a study by Krupinski et al and 0.84 vs 0.73 in a study by Barnard and Goldyne.

Logically, histological findings were more often available in the studies in which only tumors were assessed. Once again, overall analysis is impossible because several different methodologies were used. In the largest study, which analyzed 657 cases with biopsy, the validity of asynchronous teledermatology was 0.71 compared to 0.49 obtained by primary care physicians after in-person consultation. These findings indicate that the diagnosis of a teledermatologist has greater validity than an in-person assessment by a primary care physician. Unfortunately, the validity of the dermatologists’ assessment after in-person consultation (the real gold standard) was not measured. Joliffe et al, who studied the diagnosis of 144 pigmented lesions, reported a disconcerting result as they found the diagnostic accuracy of asynchronous teledermatology to be greater than that of face-to-face consultation (0.47 vs 0.43), although the difference was not statistically significant. A similar result was reported in the first study by Piccolo et al, which used digital clinical and dermoscopic images as a basis for telediagnosis. They reported an accuracy for telediagnosis of more than 0.90, higher than that obtained in face-to-face diagnosis in some cases. In 2002, however, the same authors subsequently reported a lower validity for store-and-forward teledermatology than for clinic-based consultation (0.864 vs 0.924; \( P = 0.010 \)). The authors of another similar study using dermoscopy reported excellent validity, somewhat higher for clinic-based consultation (0.911) than for asynchronous teledermatology (0.888).

The most recent studies by Moreno-Ramírez et al, which used very rigorous methodology, demonstrated the high validity of store-and-forward teleconsultation for the assessment of tumors in general and pigmented lesions in particular. Accuracy increases when teledermoscopy is used, but in view of the fact that clinical teleconsultation already has the maximum sensitivity value of 1 (if a high degree of diagnostic certainty is required from the teledermatologist), it is not clear that the improvement in specificity obtained would justify the cost of investing in dermoscopes for primary care facilities.

**Outcomes Studies**

To date, outcome studies have only evaluated intermediate results, such as the reduction in the number of unnecessary consultations, waiting time before the first intervention, and time spent on consultations. No sufficiently large studies using specific measurement instruments have been
undertaken to evaluate results such as clinical outcomes or quality of life.\(^\text{18}\)

Five studies analyzed the percentage of in-person consultations averted after a store-and-forward teledermatology consultation.\(^{26,27,42,50,82}\) The percentage ranged from 18% to 42% with a mean of 29%. Five studies evaluated the same outcome after real-time video teleconsultation, and the results ranged from 44% to 82% with a mean of 61%.\(^{72,74,76,83}\) Loane et al.,\(^{76}\) who compared store-and-forward and real-time teledermatology using the same sample group, reported that the former averted a face-to-face consultation in 31% of cases and the latter in 54%. This evidence would indicate that real-time teledermatology systems are clearly more useful than asynchronous techniques in the management of patients.

In a recent review of the Peterborough project in the UK, only 8% of 325 patients enrolled over a 51-month period did not require an in-person consultation after an asynchronous teleconsultation.\(^{85}\) In a very recent randomized study of store-and-forward teledermatology, Bowns et al.\(^{77}\) reported that only 20% of patients did not require a subsequent in-person consultation, and suggested that this technology is not suitable for this type of screening. However, in a large study by Moreno-Ramírez et al.\(^{78}\) of 2009 patients, almost 51% of referrals for possible tumors were rendered unnecessary by asynchronous teledermatology. In a Dutch postimplementation study of 503 patients who consulted their general practitioners, Knol et al.,\(^{79}\) reported that telediagnosis averted referral in 51% of the cases in which the primary care physician had originally intended to refer the patient to a specialist. These findings support the use of store-and-forward teledermatology for screening dermatological patients.

The waiting time before the first assessment by a dermatologist was shorter for asynchronous teledermatology in 3 studies: 40 vs 127 days in the study by Whited et al.;\(^{80}\) 2 vs 17 days in the study by van der Akker et al.;\(^{81}\) and 12 vs 88 days in the study by Moreno-Ramírez et al.\(^{77}\) However, in all of these studies the design may have biased the final results.\(^{12}\)

The time spent by dermatologists on an asynchronous teleconsultation was very variable. The longest time ranged from 7 minutes\(^{86}\) to 10 minutes,\(^{87}\) and the shortest from 1.5 minutes\(^{26}\) to under a minute.\(^{39}\) The results for face-to-face consultations and real-time consultations were more uniform, with between 16 and 24 minutes spent on the former,\(^{75}\) and on the order of 20 to 23 minutes on the latter.\(^{76,88}\) In a recent study that specifically analyzed the time spent by primary care physicians on a store-and-forward teledermatology consultation, the result ranged from 7 to 26 minutes, with a mean of 11 minutes. Some 41% of the time was spent on recording the clinical history and 28% on completing the electronic referral form. Initial teleconsultations and female patients took up more of the referring physician’s time.\(^{89}\) In 1 study, the referring physician spent 17.4 minutes vs 11.6 minutes to carry out a store-and-forward teledermatology consultation depending on whether or not a dermoscopic photograph was taken.\(^{71}\)

No teledermatology studies have investigated the question of second or follow-up visits or home-based telemonitoring.\(^{12}\) Only 1 study has evaluated the results of management (at 3 months)\(^{90}\); no differences in efficacy were found in terms of treatment prescribed between face-to-face consultation and store-and-forward teleconsultation. This result is particularly valuable since it came from a randomized study.

### Cost-Analysis

Two cost-analysis studies of store-and-forward teledermatology concluded that teleconsultation reduced

---

### Table 5. Cost Analysis in Teledermatology

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Analysis</th>
<th>Cost TD</th>
<th>Cost IPC</th>
<th>Teleconsultation</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zelickson, 1997(^{75})</td>
<td>Cost</td>
<td>$71.45 pp</td>
<td>$105 pp</td>
<td>ATD</td>
<td>Nursing Home</td>
</tr>
<tr>
<td>Wootton, 2000(^{72})</td>
<td>Cost-benefit</td>
<td>£132 pp</td>
<td>£49 pp</td>
<td>RTTD</td>
<td>General</td>
</tr>
<tr>
<td>Loane, 2001(^{74})</td>
<td>Cost-benefit</td>
<td>£146 pp</td>
<td>£47 pp</td>
<td>RTTD</td>
<td>Urban</td>
</tr>
<tr>
<td>Loane, 2001(^{74})</td>
<td>Cost-benefit</td>
<td>£180 pp</td>
<td>£49 pp</td>
<td>RTTD</td>
<td>Rural</td>
</tr>
<tr>
<td>Laminen, 2000(^{83})</td>
<td>Cost</td>
<td>FM18.6 tc</td>
<td>FM18 tc</td>
<td>RTTD</td>
<td>General</td>
</tr>
<tr>
<td>Bergmo, 2000(^{75})</td>
<td>Cost-minimization</td>
<td>NKr470 tc</td>
<td>NKr635 tc</td>
<td>RTTD</td>
<td>General</td>
</tr>
<tr>
<td>Chan, 2001(^{41})</td>
<td>Cost</td>
<td>HK$57 pp</td>
<td>HK$322 pp</td>
<td>RTTD</td>
<td>Nursing Home</td>
</tr>
<tr>
<td>Loane, 2001(^{73})</td>
<td>Cost-minimization</td>
<td>NZ$279 pp</td>
<td>NZ$283 pp</td>
<td>RTTD</td>
<td>Rural</td>
</tr>
<tr>
<td>Whited, 2003(^{75})</td>
<td>Cost-effectiveness</td>
<td>$36.40 pp</td>
<td>$21.40 pp</td>
<td>ATD</td>
<td>General</td>
</tr>
<tr>
<td>Armstrong, 2007(^{72})</td>
<td>Cost-minimization</td>
<td>$274/h</td>
<td>$346/h</td>
<td>RTTD</td>
<td>Rural</td>
</tr>
</tbody>
</table>

Abbreviations: ATD, asynchronous teledermatology; FM, Finnish Mark; IPC, in-person consultation; Nkr, Norwegian krone; pp, per patient; RTTD, real time teledermatology; tc, total cost; TD, teledermatology; $HK, Hong-Kong dollars; $NZ, New Zealand dollars; £, UK pounds sterling.
the cost of managing skin cancer cases and the care of patients living in institutions. In another study, store-and-forward teledermatology was shown to be more expensive than in-person consultation for the health service ($36 vs $21 per patient) but more cost-effective from a social standpoint taking into account the cost to the patient of travel and lost work time.

Initial studies of interactive consultation via a video link concluded that real-time teledermatology is more expensive than conventional clinic-based care (Table 5). However, it may be more cost effective in rural areas (particularly depending on the distance between the primary care centre and the hospital) and in the care of patients living in institutions (especially elder care and correctional facilities). It should be emphasized that the cost savings apply to the patients (travel, work time lost, etc) and not to the health care system, for which teledermatology is generally more costly than clinic-based consultation.

These studies were published at least 6 years ago, and the cost of technology in terms of personal computers and networks has decreased considerably in that period. Furthermore, all health care centers now have powerful computing equipment and telecommunications networks which they use for many different tasks. Currently, the cost of investment in equipment is falling sharply. In fact, the most recently published study reports that real-time teledermatology would be more cost-effective than in-person consultation.

### Table 6. Summary of Studies Analyzing Patient Satisfaction

<table>
<thead>
<tr>
<th>Modality</th>
<th>Author, Year</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATD</td>
<td>Pak, 1999102</td>
<td>75% would recommend TD</td>
<td>37% rated the system as poor; long wait for results</td>
</tr>
<tr>
<td></td>
<td>Kvedar, 1999103</td>
<td>42% preferred TD</td>
<td>No follow-up after consultation in 47%</td>
</tr>
<tr>
<td>RTTD</td>
<td>Nordal, 2001103</td>
<td>Overall satisfaction 93%</td>
<td>30% wanted to be able to talk to the dermatologist</td>
</tr>
<tr>
<td></td>
<td>Loane, 199897 and Gilmour, 199835</td>
<td>61% reported no disadvantages</td>
<td>14% saw the absence of palpation as a limitation</td>
</tr>
<tr>
<td></td>
<td>Artiles, 200496</td>
<td>59%-66% considered TD to be as good as in-person consultation</td>
<td>Between 13% to 18% were uncomfortable in the presence of the video camera</td>
</tr>
</tbody>
</table>

### Table 7. Summary of Studies Investigating Clinician Satisfaction (Primary Care Physicians and Dermatologists)

<table>
<thead>
<tr>
<th>Modality and Type of Clinician</th>
<th>Author, Year</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATD/PCP</td>
<td>Weinstock, 2002101</td>
<td>74% would recommend TD</td>
<td>The teleconsulting process takes up a lot of time</td>
</tr>
<tr>
<td>Pak, 1999102</td>
<td>Greater educational benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kvedar, 1999103</td>
<td>Would continue to use TD and consider it to be useful</td>
<td>Time required to complete teleconsultation form</td>
<td></td>
</tr>
<tr>
<td>Van den Akker, 200187</td>
<td>Educational benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTTD/PCP</td>
<td>Gilmour, 199895</td>
<td>75% reported an educational benefit</td>
<td>Problems with sound and image quality</td>
</tr>
<tr>
<td>Jones, 1996105</td>
<td>Benefit in CIST</td>
<td>Time spent doing the consultation</td>
<td></td>
</tr>
<tr>
<td>ATD/dermatologist</td>
<td>Pak, 1999102</td>
<td>70% of the consultations were of adequate quality</td>
<td>Less confidence in telediagnosis than clinic-based diagnosis</td>
</tr>
<tr>
<td>Van den Akker, 200187</td>
<td>Would continue to use TD</td>
<td>Time required to complete teleconsultation form</td>
<td></td>
</tr>
<tr>
<td>RTTD/dermatologist</td>
<td>Lowitt, 199892</td>
<td>98% achieved good communication with the patient</td>
<td>Less confidence in telediagnosis than clinic-based diagnosis</td>
</tr>
<tr>
<td>Nordal, 200132</td>
<td>80% found TD similar to IPC</td>
<td>Better contact with patients in IPC</td>
<td></td>
</tr>
<tr>
<td>Artiles, 200498</td>
<td>71% were satisfied with the interaction</td>
<td>48% said they would have obtained more information in IPC</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ATD, asynchronous teledermatology; RTTD, real-time teledermatology; TD, teledermatologist.
Satisfaction Studies

It should be noted that no instruments have been specifically validated for measuring the satisfaction of doctors and patients in this area.

Patients did not indicate a clear preference for either teleconsultation or conventional in-person visits (Table 6). Nor did they express any clear preference for asynchronous teledermatology or real-time interactive video consultation. The systematic reviews on this subject highlight the fact that the research in the literature indicates an acceptable degree of patient satisfaction, but go on to say that the studies undertaken have obvious methodological defects.

In 1 study, the authors make express mention of the low level of acceptance on the part of primary care physicians of a store-and-forward system in which these clinicians were responsible for collecting the clinical information, taking clinical photographs, and transmitting this data to the dermatologist. This work overload led to a situation in which only 23% of the primary care clinicians considered using teledermatology in the future. This complaint was reported in some studies but not in others.

Referring physicians usually report an educational benefit with both asynchronous and real-time teleconsultation.

Dermatologists consider both asynchronous and real-time teledermology systems to be adequate for diagnosis, although they feel less confident with teleconsultation than with in-person consultation (Table 7).

Final Analysis

Teledermatology has been shown to be feasible and reliable, but it is far from being a mature medical technology suitable for routine practice. The most studied aspect has been reliability in diagnosis and management. While reliability has been analyzed in a large number of studies, many of these have significant methodological defects. In the area of skin cancer, studies using correct methodology have demonstrated that the accuracy of teledermatology is comparable to that of in-person examination.

It is difficult to balance the conflicting needs of the implementation and the validation of teledermology systems. What is needed are studies designed to facilitate a reliable analysis of final outcomes, costs, and satisfaction with the teledermology system.

Conflicts of Interest

The authors declare no conflicts of interest.

REFERENCES

23. Phillips CM, Burke WA, Shechter A, Stone D, Balch D, Gustke S. Reliability of dermatology teleconsultations with the use of...


