Maturity of teledermatology evaluation research: a systematic literature review

N. Eminović, N.F. de Keizer, P.J.E. Bindels* and A. Hasman
Departments of Medical Informatics and *General Practice, Academic Medical Centre, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Summary

Background There is a growing interest in teledermatology in today’s clinical practice, but the maturity of the evaluation research of this technology is still unclear.

Objectives This systematic review describes the maturity of teledermatology evaluation research over time and explores what kind of teledermatology outcome measures have been evaluated.

Methods Systematic review of literature found in Medline database (1966 up to April 2006). A telemedicine evaluation strategy consisting of four consecutive research phases (parallel to drug and diagnostics evaluation research) extended with a fifth postimplementation phase was used to classify all included studies by two independent reviewers. In addition, main characteristics (store-and-forward or real-time, study design, outcome measures) were registered.

Results Three hundred and forty-five papers were systematically selected from Medline, and 244 papers were excluded. For two randomized controlled trials (RCTs), multiple papers in phase III were found. After correcting for this, 99 studies remained included (11 phase I, 72 phase II, two phase III, six phase IV, eight postimplementation phase). The number of phase II studies is the largest and still growing, while other phases are much less represented. Diagnostic accuracy was the most often used outcome measure and was found in phase I, II and IV. Store-and-forward teledermatology has been evaluated more since 2001, but most phase IV studies (RCTs, including cost aspects) are on real-time teledermatology.

Conclusions Most teledermatology evaluation studies are classified as feasibility studies (phase II). The number of phase III and IV studies remains low through the years. Compared with other specialties in telemedicine (i.e. telesurgery, telepediatrics), teledermatology seems to be a mature application. However, more evaluation studies with a focus on clinical outcomes such as preventable referrals or time to recovery are needed to prove that teledermatology indeed is a promising and cost-saving technology.

Teledermatology can be defined as the use of imaging and telecommunication technologies to provide skin services by a dermatologist to another health professional (general practitioner, nurse, other specialist) or directly to a patient. Telemedicine in general enables health professionals to communicate with each other and with their patients over a geographical or physical distance. Together with other visually oriented specialties, dermatology is one of the clinical specialties in which telemedicine is most applied.

Additional information added after online publication 5th January 2007: Preliminary results for this study were presented at the conference Medical Informatics in Europe (MIE), August 2006, Maastricht, the Netherlands.

Two types of teledermatology can be distinguished: store-and-forward (SAF) and real-time (RT) teledermatology.1 The SAF variant uses asynchronous data transfer technology (e.g. e-mail) while RT teledermatology is based on synchronous data transfer technologies (e.g. videoconferencing software). RT teledermatology requires both communicating parties (e.g. physician, dermatologist and patients) to be available at the same time, which is not necessary when using an SAF variant. This variation in teledermatology is not only technology based, but also has impact on organizational and clinical characteristics of teledermatology.

Although telemedicine and teledermatology are increasingly becoming integrated in local regular healthcare systems, there
is lack of valid scientific evidence showing positive effects, which is mainly due to the low quality of the studies. Implementation and evaluation of teledermatology seem to be performed in parallel while medical interventions are usually first fully investigated through different research phases before being implemented.

The primary aim of this review is to describe the maturity status of teledermatology evaluation research over time, for both SAF and RT, as a function of the four consecutive telemedicine evaluation phases defined by Holle and Zahlmann. These four phases are comparable with the four phases of drug research. The secondary aim of the study is to explore the outcome measures used in the different evaluation phases.

### Materials and methods

#### Search strategy

We identified English language published teledermatology studies by searching Medline database (1966 up to April 2006) using various search queries combining the key words ‘teledermatology’, ‘dermatology’, ‘telemedicine’, ‘skin’, ‘e-health’ and ‘electronic mail’. As we were only interested in original full papers reporting on the evaluation of a specific teledermatology service, we excluded literature reviews, comments, abstracts, letters and editorials. The second step included the manual screening of the selected full papers by two independent reviewers. Papers which were not about dermatology but about another specialty (e.g. radiology, pathology) and papers where the evaluation of a specific teledermatology service was not the primary aim were excluded. Papers on a telemedicine application for several specialties were included only if the results on dermatology were separately reported. Literature reviews were screened to retrieve possibly missed references. Papers from conference proceedings were excluded if a full journal paper on the same study was obtained in the selection procedure.

#### Classification of evaluation studies

Included studies were assigned independently to one of the four phases of the Holle and Zahlmann strategy. With permission of the authors, we added an extra phase: the postimplementation phase. This fifth phase should be considered as an extra category and not specifically as a continuation of the four phases. Table 1 shows the five phases which are determined by the study design, the number of participants in the study and the setting of the study.

We distinguished two main study designs: observational and intervention studies. Intervention studies were further divided into: uncontrolled trial; nonrandomized controlled trial with the same patients as control group (e.g. a patient has been seen through teledermatology first and face-to-face afterwards); nonrandomized controlled trial with other patients as control group; and randomized controlled trial (RCT).

In cases where multiple papers on one study had been published, the study was classified once in one phase. However, if multiple papers on one study were classified in different phases (for example, clinical outcomes in phase III and costs in phase IV), then the study was classified in each phase separately. When different phases of a study were described in one paper, the paper was classified only once, in the highest phase.

We predefined the following outcome measures: diagnostic accuracy, diagnostic reliability, patient satisfaction with provided care, patient and doctor satisfaction with the system, costs, preventable referrals, quality of images, learning effect in doctors and delay in treatment. When measuring diagnostic accuracy the telediagnosis is compared with a

<table>
<thead>
<tr>
<th>Phase</th>
<th>Study design</th>
<th>Usual participants</th>
<th>Main criteria used by the reviewers</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Exploratory, small intervention</td>
<td>Researchers and project members, simulating</td>
<td>Experimental setting, i.e. teleconsultation did not result in a diagnosis or treatment plan</td>
</tr>
<tr>
<td></td>
<td>studies</td>
<td>patients</td>
<td>Field or experimental studies with potential users</td>
</tr>
<tr>
<td>II</td>
<td>Feasibility studies, can be controlled</td>
<td>Specially trained and highly motivated</td>
<td>RCT on clinical outcomes</td>
</tr>
<tr>
<td></td>
<td>intervention studies, but rarely</td>
<td>potential users, real or simulated patients</td>
<td></td>
</tr>
<tr>
<td></td>
<td>involve a separate control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>RCT</td>
<td>Unselected sample of users with minimal training</td>
<td>Costs are compared between conventional dermatological care and teledermatology and/or with any kind</td>
</tr>
<tr>
<td>IV</td>
<td>(Simulation) cost studies using the</td>
<td>real patients</td>
<td>of benefits resulting from an RCT</td>
</tr>
<tr>
<td></td>
<td>benefits from an RCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postimplementation</td>
<td>Teledermatology part of the regular</td>
<td>Actual users of teledermatology</td>
<td>Telemedicine is fully integrated in the regular care</td>
</tr>
<tr>
<td></td>
<td>care, observational study</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RCT, randomized controlled trial.

© 2006 The Authors

gold standard diagnosis (face-to-face diagnosis, biopsy etc.). Diagnostic reliability is diagnostic agreement between different examiners who determined a diagnosis using the same or a different, but not gold standard, diagnostic method. Patient satisfaction with the teledermatology system refers to how the patients experience a new technology such as teledermatology, while patient satisfaction with care refers to how the patients feel about the health care delivered in general.

All discrepancies in classifying the selected papers between the reviewers were solved by consensus. Interobserver reliability regarding the classification into phases was calculated to determine the validity of the classification procedure.

Results

The Medline search resulted in 345 unique references. Of these studies, 181 studies were included and 164 excluded (see Fig. 1). Forty-five papers were about another specialty (radiology, pathology) or about all kinds of medical conditions including dermatology as a subset but without reporting separate results. After reading the full papers, 80 papers were excluded. Most of these papers did not report on an evaluation of a specific teledermatology service ($n = 46$). Five studies needed to be excluded as they were in some way duplicate publications: conference proceedings and a full version paper. Finally, 101 papers were included for the classification into five phases. It appeared that in two situations two related papers on the same RCTs were classified in the same phase, phase III. Following the rules, these studies were classified once, making the total number of classified studies as 99.

The two reviewers agreed about the classification of the studies into phases in 84% of cases ($\kappa = 0.7$).

Phase trends

In 1996 two teledermatology evaluation studies were published. These first teledermatology studies were a phase I study and phase II study (Fig. 2). Phase I studies have been published sporadically with a total of 11 studies up to 2006. Seventy-two studies were assigned to phase II. In 1998, 10 phase II studies were published while only one phase I and one postimplementation study were published. The same high number of phase II studies was found again in 2003 and 2005.

The number of phase III studies was limited to two, published in 2002 and 2004.10,11 Phase IV and the postimplementation phase were represented by six and eight studies, respectively. In 2000 three cost studies were published12–14 and another two were published in 2001.15,16 The last published phase IV study dates from 2003.17

Two large groups of trials were found: U.K. Multicentre Teledermatology Trials14,18–21 and Health Waikato Teledermatology Trials13,12,23. These trials investigated RT
teledermatology and covered different phases. The U.K. trials covered phases I–IV, while the Waikato trials covered phases II–IV.

Outcome measures and study design

The diagnostic accuracy was the most common outcome measure used in 53 studies (Table 2). In seven of 11 phase I studies, the quality of images was investigated. The two phase III RCTs investigated delay in treatment and preventable referrals as clinical outcome measures, but also included patient and doctor satisfaction with the system. Central themes of the eight postimplementation phase studies were patient and doctor satisfaction with the teledermatology system.

In total, there were 43 intervention studies controlled by the same group of patients. An uncontrolled study design was applied in 30 studies, mainly in phase II (21 of 72), but also in phase I (six of 11). We found four RCTs in phase II which were small and not focusing on clinical outcomes and could therefore not be classified as phase III or IV.24–27

Store-and-forward vs. real-time teledermatology

The majority of the included studies (63%) investigated the SAF variant of teledermatology (Fig. 3). The RT variant was investigated in 29% of the studies, with the majority performed before 2001. Since 2001 more SAF applications were the subject of evaluation compared with RT teledermatology applications. The combination of the two technologies was evaluated twice.19,28 In six studies it was not clear what kind of technology was used. Three of these six studies were phase I studies in which technical aspects and quality of images were evaluated.29–31 The other three studies were two recent phase II13,33 and one postimplementation phase study.34

SAF technology was investigated in all phase III studies, while four of six phase IV studies evaluated RT12,13,15,16 and one SAF teledermatology.17

Table 2 Outcome measures and study designs in five phases

<table>
<thead>
<tr>
<th>Phase</th>
<th>n (SAF, RT, both)</th>
<th>Outcome measures (number of studies)</th>
<th>Study design (number of studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I8,21,29–31,38–43</td>
<td>11 (5, 3, 0)</td>
<td>Diagnostic accuracy (5)</td>
<td>Observational (2)</td>
</tr>
<tr>
<td></td>
<td>3 unknown</td>
<td>Diagnostic reliability (3)</td>
<td>Intervention controlled by the same patients (3)</td>
</tr>
<tr>
<td></td>
<td>SAF or RT</td>
<td>Quality of images (7)</td>
<td>Uncontrolled (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Doctor satisfaction with the system (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preventable referrals (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confidence in the findings (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Technical aspects (2)</td>
<td></td>
</tr>
<tr>
<td>II9,18–20,22–28,32,33,44–102</td>
<td>72 (47, 18, 1)</td>
<td>Diagnostic accuracy (47)</td>
<td>RCT (4)</td>
</tr>
<tr>
<td></td>
<td>2 unknown</td>
<td>Diagnostic reliability (17)</td>
<td>Intervention controlled by other patients (7)</td>
</tr>
<tr>
<td></td>
<td>SAF or RT</td>
<td>Quality of images (12)</td>
<td>Intervention controlled by the same patients (40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patient satisfaction with the system (16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Doctor satisfaction with the system (11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patient satisfaction with care (5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Costs (5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preventable referrals (12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Learning effect GPs (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confidence in the findings (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time needed for consultation (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Technical aspects (3)</td>
<td></td>
</tr>
<tr>
<td>III10,11</td>
<td>2 (2, 0, 0)</td>
<td>Patient satisfaction with the system (2)</td>
<td>RCT (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Doctor satisfaction with the system (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patient satisfaction with care (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preventable referrals (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Delay in treatment (2)</td>
<td></td>
</tr>
<tr>
<td>IV12–17</td>
<td>6 (1, 4, 1)</td>
<td>Diagnostic accuracy (1)</td>
<td>RCT (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Costs (6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preventable referrals (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time needed for consultation (1)</td>
<td></td>
</tr>
<tr>
<td>Postimplementation4,103–109</td>
<td>8 (4, 3, 0)</td>
<td>Patient satisfaction with the system (2)</td>
<td>Observational (5)</td>
</tr>
<tr>
<td></td>
<td>1 unknown</td>
<td>Doctor satisfaction with the system (2)</td>
<td>Uncontrolled (3)</td>
</tr>
<tr>
<td></td>
<td>SAF or RT</td>
<td>Costs (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Delay in treatment (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time needed for consultation (1)</td>
<td></td>
</tr>
</tbody>
</table>

SAF, store-and-forward; RT, real-time; GP, general practitioner; RCT, randomized controlled trial.

© 2006 The Authors
Fig 3. Cumulative number of store-and-forward (SAF) vs. real time (RT) teledermatology evaluation studies in the past 10 years.

Discussion

Teledermatology has been the subject of evaluation studies for 10 years now. The results of our review show that evaluation of teledermatology takes place in different research phases. Despite a growing implementation of teledermatology in today’s clinical practice, there are no clear trends towards phase III and IV studies in more recent years in comparison with the first years of teledermatology evaluation.

The number of phase II studies grew continuously over the past 10 years and is still growing. Phase III, IV and V studies are much less represented. There are different possible explanations for this. Phase I studies, usually reporting on technical issues such as, for example, quality of images, are often not published separately as scientific journal papers (possible publication bias). Technical evaluations are also more likely to be published in technical journals which are not covered by the Medline database.

Only two RCTs could be classified in phase III and six in phase IV. Performing high-level evidence studies such as RCTs in teledermatology is often seen as difficult as teledermatology is a complex intervention depending on local variations in administration and organisation.35,36 Characteristics of teledermatology such as its relatively low risk, low investment costs and the possibility of using it in addition to regular care are likely to speed up its implementation despite the lack of scientific research on the pros and cons. However, detailed and sophisticated evaluation of any new technology requires many resources and should be performed only after some basic but important evaluation steps have been carried out in phase I and II. Furthermore, for certain nonclinical outcome measures of teledermatology (e.g. diagnostic accuracy), a phase II study should provide sufficiently strong evidence and a phase III or IV study is not expected. These issues are likely to reduce the push to perform RCTs in teledermatology.

Postimplementation studies are, like phase I studies, not often published in scientific journals. Commercial teledermatology suppliers probably perform this kind of evaluation frequently, but mainly for their own use without academic ambitions. In our review we obtained only Medline indexed literature. Further in-depth analysis of phase I or postimplementation phase studies would require obtaining literature databases of technical journals and nonpublicly available reports.

More detailed and sophisticated evaluation of any new technology is expected to be carried out after some basic evaluation steps have been performed, such as in phase I and II. As no or limited scientific evidence about the impact of teledermatology on clinical outcome has been demonstrated,4 more detailed evaluation studies (phase III and IV) are needed.

Diagnostic accuracy is the most often evaluated outcome measure in the teledermatology studies included in our review. This result was not surprising as this outcome measure is strongly related to the visual aspect of teledermatology. With the changing technology (i.e. different cameras with different possibilities and settings) and large variations in teledermatology settings, this outcome is likely to be studied repeatedly in most teledermatology evaluations. However, it is regrettable that important clinical outcomes such as preventable referrals and delay in treatment are not more often investigated, especially in phase III and IV studies.

Earlier studies showed that RT teledermatology was more clinically efficient than the SAF variant.14 However, SAF teledermatology is less expensive and under certain circumstances more cost effective, making it more attractive than the RT variant. This increasing interest in SAF has also been shown in our review as RT was especially popular in the early evaluations and SAF has been much more popular over the past 5 years. However, most phase IV studies that we found were about RT. Therefore we support Whited’s conclusion that there is a lack of economic evaluation of SAF teledermatology.7

An earlier study concluded that teledermatology was maturing because of the reasonably good quality and quantity of the published literature when compared with other specialties (i.e. telesurgery, telepaediatrics).37 Hailey et al.4 found for teledermatology the strongest indication for benefits of telemicine applications. Although these conclusions might be true when compared with other telemicine applications, our review shows that there is no clear trend in maturity of teledermatology evaluation research over the past 10 years as there is a lack of phase III and IV studies. At this moment, we found no indications for the maturation of teledermatology evaluation research or that the number of phase III and phase IV studies is likely to increase in the near future. The timing of our study might be too early to measure an increase of phase III and phase IV studies as these

© 2006 The Authors
studies are often time consuming in both performing and reporting. Although teledermatology seems to be a valuable application, we believe that there is a need for more phase III and IV studies to provide a high level of evidence on positive clinical outcomes and cost effectiveness of teledermatology prior to its implementation and wide use in clinical practice.

References

Maturity of teledermatology evaluation research, N. Eminović et al.


© 2006 The Authors