

Diagnostic Reliability of In-Person Versus Remote Dermatology: A Meta-Analysis

Mrigendra Bastola, MD, MS, Craig Locatis, PhD,
and Paul Fontelo, MD, MPH

National Library of Medicine, National Institutes of Health,
Bethesda, Maryland, USA.

Abstract

Background: Studies comparing teledermatology with in-person dermatologists report wide variations in diagnostic agreement. Teledermatology studies should have two independent in-person consultations establishing a baseline for comparing diagnoses made face-to-face and those made remotely.

Objective: To perform a meta-analysis of comparison studies having two in-person dermatologists and at least one remote dermatologist examining the same patients to determine the overall preponderance of agreement.

Method: Studies having two in-person diagnosticians were identified from previous teledermatology research reviews and independent searches of PubMed and other databases. Data from six studies identified were meta-analyzed.

Results: Some studies showed high levels of diagnostic concordance, while others did not. Meta-analysis revealed that concordance rates reported in the teledermatology and clinical (in-person) consultations were significantly different (odds ratio=0.55 [Mantel-Haenszel, fixed effect model, 95% confidence interval=0.42–0.72], $\chi^2=11.87$, $p < 0.05$, $I^2=58\%$). Overall results showed that in-person primary diagnoses are significantly more concordant than remote. The results also suggest that diagnoses made in-person and teledermatology were marginally but significantly different than remote.

Conclusion: Although the results of this study suggest teledermatology diagnoses are less reliable than those in-person, there are still valid reasons for using teledermatology to improve access, reduce costs, and triage patients to determine those warranting further in-person consultation and/or laboratory tests. More caution should be exercised in teledermatology when diagnoses involve risky skin conditions. There is evidence that this happens in practice.

Keywords: telemedicine, telehealth, teledermatology, e-health, dermatology

Background

Teledermatology consultations usually are either live interactive, where dermatologists examine patients in real time by videoconference or store-and-forward, where photos and histories taken by a general practitioner or other trained personnel are reviewed by a dermatologist asynchronously. Most teledermatology studies compare diagnoses between a single dermatologist in a clinic and a remote one or, in some cases, a dermatologist making a diagnosis in-person and then a rediagnosis after a time-lapse, from pictures and histories taken during the initial examination. This meta-analysis includes data from studies using both types of teledermatology interventions. The study aims to meta-analyze data from well-conducted studies, where patients were independently evaluated by at least two dermatologists in-person.

Early telemedicine reviews identified a problem with studies having single clinic and remote diagnosticians; namely, when there is disagreement, it is not possible to determine who is correct. The reliability problem is resolvable by having two or more in-person dermatologists make independent diagnoses of the same patient to establish a concordance baseline for comparing remote diagnoses.^{1,2} The reliability of diagnoses made by telemedicine depends on their agreement with those made in-person since remote diagnoses need only be as good as, not better than those face-to-face.^{1,2} Teledermatology research reviews report concordance between diagnoses made remotely and those made in-person ranging from 46% to 99%.^{3–5}

One 2009 teledermatology research review identified 12 reliability studies with multiple dermatologists,⁴ but 8 of these studies had multiple remote dermatologists.^{6–13} Only four studies cited in the review had two in-person dermatologists,^{14–17} including a pilot study¹⁴ for a more extensive follow-up.¹⁵ The reliability issue is compounded by variations in the definition of the term “partial agreement.” While all study authors agreed that complete agreement is an exact match of primary diagnoses, there are varied definitions of a “partial” agreement. Partial agreement has been defined as the primary diagnosis of one examiner being in the differential of another,¹⁵ any shared diagnoses in examiner differentials,¹⁶ or differentials matching in all diagnoses, even if their order does not.¹⁸

The pilot study cited in the 2009 review involved 12 patients referred for suspected skin cancer who were examined by 2 dermatologists in clinic and 2 different dermatologists reviewing patient digital images. The clinical diagnosticians' agreement was 80% for primary and 90% for partial, while their primary agreement with digital diagnosticians ranged from 80% to 39% for primary and from 100% to 90% for partial. The digital diagnosticians' agreement with each other was 46% for primary and 92% for partial.¹⁴ A more extensive follow-up study had 129 patients who were seen by 2 different dermatologists in-person and who had their digital images reviewed by 3 different dermatologists. Kappa correlations were calculated to determine agreement among clinical and digital examiners, with the value 1 representing complete agreement. The kappa values for clinical and digital primary diagnoses were both 0.68, while those for partial were ≥ 0.60 .¹⁵

A third study cited in the 2009 review had 892 patients who were assigned to conventional clinical examinations or review by store-and-forward teledermatology. Primary diagnostic agreement was 55% for store-and-forward examiners and 79% for the two clinical consultants. The agreement levels were statistically significantly different.¹⁷ A fourth study cited had 60 patients with skin problems who were seen in-person and remotely by videoconference. After half the patients were seen, the two dermatologists switched being in-person and remote examiners. An additional 36 patients were examined independently in-person by both dermatologists as a control. There was a 94% complete agreement for the control group with a partial agreement for all of the remaining 6%. Complete agreement between in-person and teledermatologists was 78%, with partial agreement for the remaining 21% and no agreement for the remaining 1%.¹⁷

Two additional two in-person dermatologist studies have been conducted since the 2009 review. The first had 174 patients who were examined by two dermatologists face-to-face

and two other dermatologists reviewing digital images and clinical histories. Agreement between two in-person dermatologists for the primary diagnosis was 83.3% compared with agreement of 81.0% for two dermatologists reviewing images and histories. Agreement between in-person dermatologists and dermatologists viewing images ranged from 78.2% to 83.9%.¹⁸ The second study compared examinations of 214 patients seen in-person by two dermatologists (an attending and resident) making a consensus diagnosis with diagnoses made of the same patients by other dermatologists using uncompressed and compressed high-definition video and store-and-forward methods. A subset of 134 patients also were evaluated independently by the two in-person dermatologists to establish a baseline. In-person attending and resident kappa values with the consensus were 0.91 for primary agreement and 1.0 for partial agreement. Comparable kappa values were 0.76 for primary agreement and 0.87 for partial agreement with the store-and-forward method, 0.76 for primary agreement and 0.89 for partial agreement with uncompressed video, and 0.72 for primary agreement and 0.88 for partial agreement with compressed video.¹⁹

Methods

This study followed the PRISMA guidelines for meta-analysis.²⁰ Teledermatology research reviews through 2019 were analyzed to identify studies having two in-person consultations. Since a 2009 review identified the most, it was used as a starting point.⁴ Two of the authors independently searched PubMed registries from 2008 to 2019 looking for appropriate search terms in either article titles or abstracts. Search terms included were teledermatology, in-person dermatology, remote dermatology, telehealth, e-health, concordance, reliability, effectiveness, and digital. If titles suggested they focused on accuracy, reliability, or agreement, the abstracts from these studies' abstracts were evaluated to

Table 1. Characteristics of Included Studies

| STUDY | YEAR | CASE TYPES SELECTED | TELEDERMATOLOGY ENCOUNTERS | IN-PERSON ENCOUNTERS | NUMBER OF TELEDERMATOLOGISTS | NUMBER OF CLINICIANS |
|-------------------------------|------|---------------------|----------------------------|----------------------|------------------------------|----------------------|
| Browns et al. ¹⁷ | 2006 | Multiple | 92 | 73 | 2 | 2 |
| Leshner et al. ¹⁶ | 1998 | Multiple | 68 | 47 | 2 | 2 |
| Marchell et al. ¹⁹ | 2017 | Multiple | 162 | 162 | 2 | 2 |
| Ribas et al. ¹⁸ | 2010 | Multiple | 174 | 174 | 2 | 2 |
| Whited et al. ¹⁴ | 1998 | Cancer | 46 | 10 | 2 | 2 |
| Whited et al. ¹⁵ | 1999 | Multiple | 168 | 168 | 3 | 2 |

determine if these factors were assessed and if there were at least two in-person consultations. If there was still ambiguity, the entire article was reviewed. The independent searches by two authors identified the exact same two studies for inclusion, bringing the total studies for meta-analysis to 6. All six studies identified from the reviews and the searches have been described above and are listed with their key characteristics in *Table 1*. One author independently searched other databases, including some having articles in international languages, and found no additional studies.

Given the varied definitions of partial agreement, the meta-analysis focused on primary agreement only. Review Manager (version 5.3) was used for data sorting and forest plot generation. Cochrane risk of bias tool was utilized to assess risks. Unit of analysis was identified, and the corresponding authors were contacted, if needed, in cases when data needed to be clarified or data appeared to be missing. The Mantel-Haenszel (M-H) method assuming a fixed effect model was used for the meta-analysis due to the small number of the included studies. No subgroup or sensitivity analyses were done.

Results

A forest plot was created based on study inclusion criteria for reported diagnostic agreement events of both in-person and remote dermatologists. The results of the meta-analysis revealed that the concordance rates reported in the teledermatological and clinical (in-person) consultations were significantly different (odds ratio = 0.55 [M-H, fixed effect model, 95% confidence interval = 0.42-0.72], $\chi^2 = 11.87, p < 0.05, I^2 = 58\%$) (*Fig. 1*). The I^2 value of 58% indicates the presence of a moderate degree of heterogeneity in the analysis.

Discussion

The need for having a two in-person agreement baseline for comparing remote diagnoses goes back to some early telemedicine reviews.^{1,2} It was surprising that only six reliability studies could be identified meeting this requirement over an

almost 20-year period. While several of these studies found significant levels of diagnostic concordance between the teledermatologists and in-person dermatologists, the meta-analysis indicates that diagnostic concordance in teledermatology is significantly, but marginally, lower than that for diagnoses made in in-person and that studies having only a single teledermatologist had larger concordance differences. Results suggest that caution should be exercised when using teledermatology. There is evidence that this is indeed the case. One study included in the meta-analysis¹⁹ and an earlier correlation study²¹ showed that more biopsies are ordered when teledermatology is used, indicating precautions are taken. There is always the option of still seeing patients in-person when there is diagnostic uncertainty or patients are suspected of having conditions of high risk.

Reliability is only one of the many indicators of telemedicine's efficacy. Teledermatology can be used to improve access and reduce costs and time to render care. It can be used as a substitute for face-to-face examinations when patients present with conditions having low risks and as a method to triage patients for in-person consultations and biopsy when riskier diagnoses are possible.²²

One limitation of this meta-analysis is the small number of studies on which it is based. Moreover, this study did not examine diagnostic accuracy, usually defined as comparing diagnoses to histopathology, because only a subset of patients in the studies needed biopsies and this information usually was not reported. Although not a limitation of this study, other potential sources of bias include the discrepancies in the quality of the images used for analysis, which is presumably the same within a study but could vary between studies occurring at different places and times using different technologies. Also, this meta-analysis showed a moderate level of heterogeneity, which could be related to many factors, such as the differences in demographics, differences in consultant experience, variable skin tones and colors, and differences in imaging methods.

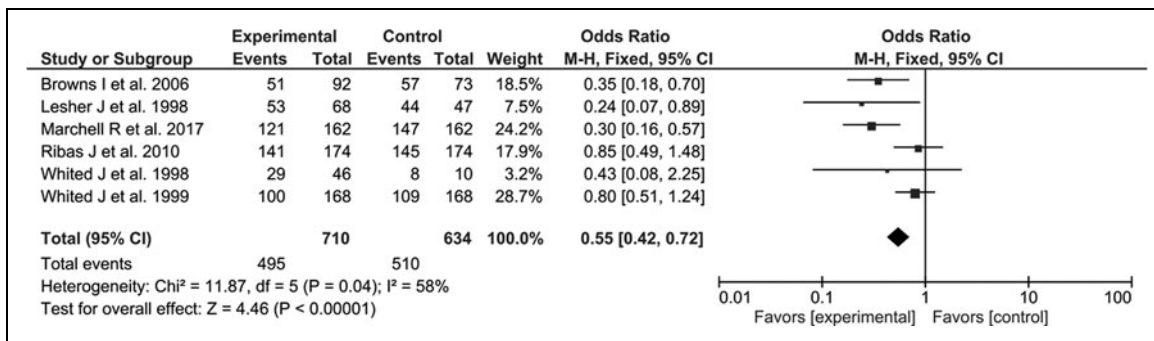


Fig. 1. Forest plot of the included studies.

Conclusions

The meta-analysis indicates that the agreement of remote diagnoses is significantly, but marginally, lower than the agreement between diagnoses made in-person. Although less reliable than consultations made face-to-face, there are still valid reasons for providing teledermatology services, not the least of which are providing care more rapidly to those who might not get it otherwise and as a way of screening patients who might safely be provided care at a distance and those who might need to be seen in-person or need additional tests.

Acknowledgment

The authors are grateful for the thoughtful review of the study's data analysis provided by biostatistician Dr. Seo Hyon Baik.

Disclaimer

The views and opinions of the authors herein do not necessarily state or reflect those of the National Library of Medicine, the National Institutes of Health, or the U.S. Department of Health and Human Services.

Disclosure Statement

No competing financial interests exist.

Funding Information

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), the National Library of Medicine (NLM), and the Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

- Hersch W, Wallace J, Patterson P, Shapiro S, Kraemer D, Eilers G, Chan B, Greenlick M, Helfand M. *Telemedicine for the Medicare population. Evidence Reports/Technology Assessments, No. 24*. Report No. 01-E012. Rockville, MD: Agency for Healthcare Research and Quality, 2001.
- Hersch W, Hickam D, Severance S, Dana T, Krages K, Helfand M. *Telemedicine for the Medicare population: Update. Evidence Reports/Technology Assessments, No. 131*. Report No. 06-E007 Rockville, MD: Agency for Healthcare Research and Quality, 2006.
- Johnson M, Armstrong A. Technologies in dermatology: Teledermatology review. *G Ital Dermatol Venereol* 2011;146:143–153.
- Levin Y, Warshaw E. Teledermatology: A review of reliability and accuracy of diagnosis and management. *Dermatol Clin* 2009;27:163–176.
- Warshaw E, Hillman Y, Greer N, Hagel E, MacDonald R, Rutks I, Wilt T. Teledermatology for diagnosis and management of skin conditions: A systematic review. *J Am Acad Dermatol* 2011;64:759–772.
- Baba M, Seckin D, Kapdagli S. A comparison of teledermatology using store-and-forward methodology alone, and in combination with Web camera videoconferencing. *J Telemed Telecare* 2005;11:354–360.
- Krupinski E, LeSueur B, Ellsworth L, Levine N, Hansen R, Silvis N, Sarantopoulos P, Hite P, Wurzel J, Weinstein R, Lopez A. Diagnostic accuracy and image quality using a digital camera for teledermatology. *Telemed J* 1999;5:257–263.
- Oztas M, Calikoglu E, Baz K, Birol A, Onder M, Calikoglu T, Kitapci M. Reliability of Web-based teledermatology consultations. *J Telemed Telecare* 2004;10:25–28.
- Lim A, Egerton I, See A, Shumack S. Accuracy and reliability of store-and-forward teledermatology: Preliminary results from the St. George Teledermatology Project. *Australas J Dermatol* 2001;42:247–251.
- Monero-Ramirez D, Ferrandiz L, Nieto-Garcia A, Carrasco R, Moreno-Alvarez P, Galdeano R, Bidegain E, Rios-Martin J, Camacho F. Store-and-forward teledermatology in skin cancer triage: Experience and evaluation of 2009 teleconsultations. *Arch Dermatol* 2007;143 479–484.
- Moreno-Ramirez D, Ferrandiz L, Perez-Bernal A, Carrasco-Duran R, Rios-Martin J, Camacho F. Teledermatology as a filtering system in pigmented lesion clinics. *J Telemed Telecare* 2005;11:298–303.
- Massone C, Hofmaann-Wellenhof R, Ahlgrimm-Siess V, Gabler G, Ebner C, Soyer H. Melanoma screening with cellular phones. *PLoS One* 2007;2:e483.
- Piccolo D, Soyer H, Chimenti S, Argenziano G, Bartenjev I, Hoffmann-Wellenhof R, Marchetti R, Oguchi S, Pagnanelli G, Pizzichetta M, Saida T, Salvemini I, Tanaka M, Wolf I, Zgavec B, Peris K. Diagnosis and categorization of acral melanocytic lesions using teledermoscopy. *J Telemed Telecare* 2004;10:346–350.
- Whited J, Mills B, Hall R, Drugge R, Grichnik J, Simel D. A pilot trial of digital imaging in skin cancer. *J Telemed Telecare* 1998;4:108–112.
- Whited J, Hall R, Simel D, Foy M, Stechuchak K, Drugge R, Grichnik J, Myers S, Horner R. Reliability and accuracy of dermatologists' clinic-based and digital image consultations. *J Am Acad Dermatol* 1999;41(5 Pt 1):693–702.
- Leshner J, Davis L, Gourdin F, English D, Thompson W. Telemedicine evaluation of cutaneous diseases: A blinded comparative study. *J Am Acad Dermatol* 1998;38:27–31.
- Browns I, Collins K, Walters S, McDonagh A. Telemedicine in dermatology: A randomized control trial. *Health Technol Assess* 2006;10:1–39.
- Ribas J, da Graca Souza Cunha M, Pedro Mendes Schettini A, da Roch Ribas CB. Agreement between dermatological diagnoses made by live examination compared to analysis of digital images. *An Bras Dermatol* 2010;85:441–447.
- Marchell R, Locatis C, Burges G, Maissiak R, Liu W, Ackerman M. Comparing high definition live interactive and store-and-forward consultations to in-person examinations. *Telemed J E Health* 2017;23:213–218.
- PRISMA (2015). PRISMA checklist, PRISMA (2015). Available at <http://prisma-statement.org/documents/PRISMA%202009%20checklist.pdf> (last accessed February 5, 2020).
- Pak H, Harden D, Cruess D, Welch M, Poropatich R. Teledermatology: An intraobserver diagnostic correlation study, Part II. *Cutis* 2003;71:476–480.
- Lowitt MH, Kessler II, Kauffman CL, Hooper FJ, Siegel E, Burnett JW. Teledermatology and in-person examinations: A comparison of patient and physician perceptions and diagnostic agreement. *Arch Dermatol* 1998;134:471–476.

Address correspondence to:
Mrigendra M. Bastola, MD, MS
 National Library of Medicine
 National Institutes of Health
 Bethesda, MD 20894
 USA

E-mail: mrigendra.bastola@nih.gov

Received: February 5, 2020

Accepted: May 13, 2020

Online Publication Date: June 23, 2020