Reliability and validity of mobile teledermatology in HIV positive patients in Botswana: a pilot study

Rahat S Azfar, M.D., M.S.C.E.[1,2], **Robert A. Lee, M.D., Ph.D.**[3], **Leslie Castelo-Soccio, M.D. Ph.D.**[1], **Martin S. Greenberg, D.D.S.**[4], **Warren B Bilker, Ph.D.**[2], **Joel M Gelfand, M.D., M.S.C.E.**[1,2], and **Carrie L Kovarik, M.D.**[1]

[1]Department of Dermatology, University of Pennsylvania, Philadelphia, USA

[2]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, USA

[3]Department of Medicine, Division of Dermatology, University of California San Diego, San Diego USA

[4]Department of Oral Medicine, School of Dental Medicine University of Pennsylvania, Philadelphia, USA

Abstract

Importance—Mobile teledermatology may increase access to care.

Objective—To determine if mobile teledermatology in HIV positive patients in Gaborone, Botswana was reliable and produced valid consultations in comparison to face-to-face dermatology consultations.

Design—Cross-sectional study

Setting—Outpatient clinics and public inpatient settings in Botswana.

Participants—76 HIV positive patients aged 18 years and up with a skin or mucosal complaint that had not been previously evaluated by a dermatologist.

Main Outcome(s) and Measure(s)—We calculated Cohen's kappa coefficient for diagnosis, diagnostic category and management for test-retest reliability and for inter-rater reliability. We also calculated sensitivity and specificity for each diagnosis.

Results—Cohen's kappa for test-retest reliability ranged from 0.47 (95% CI 0.35-0.59) to 0.78 (95% CI 0.67-0.88) for the primary diagnosis, 0.29 (95% CI 0.18-0.42) to 0.73 (95% CI

0.61-0.84) for diagnostic category, and 0.17 (95% CI -0.01-0.36) to 0.54 (95% CI 0.38-0.70) for management. Cohen's kappa for inter-rater reliability ranged from 0.41 (95% CI 0.31-0.52) to 0.51 (95% CI 0.41-0.61) for the primary diagnosis, from 0.22 (95% CI 0.14-0.31) to 0.43 (95% CI 0.34-0.53) for the diagnostic category for the primary diagnosis and from 0.08 (95% CI 0.02-0.15) to 0.12 (95% CI 0.01-0.23) for management. Sensitivity and specificity for the top ten diagnoses ranged from 0 to 0.88 and from 0.84 to 1 respectively.

Conclusions and Relevance—Our results suggest that while the use of mobile teledermatology technology in HIV-positive patients in Botswana has significant potential for improving access to care, additional work is needed to improve reliability and validity of this technology on a larger scale in this population.

Keywords

Validation study; Mobile Teledermatology; HIV

Introduction

Background

In many parts of the world, particularly in sub-Saharan Africa, there is a severe shortage of dermatologic specialists.[1] Dermatologic care is often provided by clinicians and rural health workers who have limited training in dermatology.[2] This shortage is more acutely felt in the HIV positive community in these regions, as there is an increased burden of both prevalence and severity of skin and mucosal disease in this group in comparison to the immunocompetent population. In addition, the presence of several particular mucocutaneous conditions may also affect HIV management.[3,4,5]

While traditional store-and-forward teledermatology offers a method for increasing access to skin specialists in these regions, issues with limited computer connectivity often arise. Mobile teledermatology utilizes cellular phone networks, which are more stable and accessible, to perform store-and-forward teledermatology consults.[6,7] While several studies have evaluated diagnostic agreement, relatively few have investigated the reliability and validity of mobile teledermatology in comparison to the gold standard of face-to-face evaluation by a dermatologist.[6,8,9,10] Moreover, to our knowledge, this technology has not been tested in the field in sub-Saharan African among HIV positive patients.

Objective

We sought to determine if the use of mobile teledermatology technology in HIV positive patients in Gaborone, Botswana was reliable and produced valid consultations when compared to face-to-face dermatology consultations. We hypothesized that health care workers could transmit clinical information and photos through the mobile phone that would allow reliable and valid remote dermatologic consultations that were similar in quality to in person consultations.

Methods

Study Design and Setting

We conducted a cross-sectional pilot study of adult patients with HIV and mucocutaneous complaints in Botswana. The study was approved by the Institutional Review Boards at the University of Pennsylvania (Protocol #809728, approved 4/17/09), Princess Marina Hospital, and the Botswana Ministry of Health (Protocol #HRDCOO4, approved 7/7/09).

Participants

The study was conducted in consecutively recruited HIV positive patients who were at least 18 years of age and presented with a skin or mucosal complaint that had not been previously evaluated by a dermatologist. The patients were recruited from the medical and oncology wards, the dermatology clinic, and the infectious disease clinic (IDCC) at Princess Marina Hospital in Gaborone, Botswana; from Independence Surgery Center, a private primary care clinic in Gaborone, Botswana; and from the outpatient clinics and medical wards at Athlone Hospital in Lobatse, Botswana over a 5 week period from August through September 2009.

Data sources/measurement and bias

All patients received a face-to-face clinical evaluation by a U.S. based board-certified dermatologist with clinical experience in Botswana. At the end of their clinical encounter with the dermatologist, patients were asked to participate in a mobile telephone encounter for the purpose of the study. A Setswana-speaking nurse obtained consent and clarified any patient questions. Enrolled patients received 30 pula (4 US Dollars) compensation to cover the cost of their travel at the end of the mobile encounter. The face-to-face dermatologist completed a separate de-identified clinical evaluation form for purposes of study data collection for each enrolled patient. (eFigure 1) This evaluation was used as the gold standard for comparative purposes. Patients who consented to the mobile telephone encounter were then seen by the nurse interviewer who collected their data and forwarded it for study purposes for mobile teledermatology evaluation. To simulate a typical hypothetical setting in which mobile teledermatology may be used, the nurse was trained in using the phone software for history taking and medical photography a few days prior to beginning the study, but worked independently from the face-to-face dermatologist to collect data, and had no previous training or experience in dermatology. Data were collected without any personally identifying information and forwarded directly from the Samsung Soul SGH-U900 mobile phone with 5 megapixel camera to a secure password protected teledermatology evaluation website. (eFigures 2, 3, 4) After initial data collection was completed, mobile evaluations were completed by 3 U.S. based board certified dermatologists and one board certified oral medicine specialist. (eFigure 5) Each of these evaluators had varied levels of clinical experience working in the sub-Saharan HIV-positive population or similar populations. The oral medicine specialist evaluated only those cases which included oral pathology. Mobile evaluations were not used to guide clinical decision-making and were solely conducted for study purposes. Each patient could have had multiple diagnoses; for each diagnosis, the evaluators were asked to provide a ranked differential diagnosis when they thought it was appropriate, a diagnostic category for their primary differential diagnosis (i.e., bacterial infection, neoplasm, papulosquamous inflammation,

etc.), and their recommendations for management (treat, test, test and treat, refer for face-to-face evaluation). In order to assess test-retest reliability, several months after their initial mobile evaluations were completed, the mobile evaluators were given the cases again, without access to their previous responses. The methods and forms were piloted in the dermatology clinic at Princess Marina Hospital and with the mobile teledermatologists in the U.S. in the months preceding the beginning of the study.

Variables and statistical methods

We calculated descriptive statistics for the overall cohort as well as inter-rater reliability and test-retest reliability for each of the main outcomes and sensitivity and specificity for each diagnosis using Stata 10.1. For the reliability analyses, our main outcomes were Cohen's kappa coefficient for: diagnosis, diagnostic category and management. The findings of the face-to-face dermatologist were considered the gold-standard for the purposes of determining inter-rater reliability and validity of the diagnoses. The significance level (alpha) was set at 5% for all hypothesis tests.

Study size

Study size was determined based on the anticipated number of patients needed to achieve 80% power to conduct our primary analysis; we estimated that we needed at least 108 patients with a single mucocutaneous problem.

Results

Participants

Patient characteristics have previously been described.[11] Due to the loss of over one week of allocated study time to recently placed government regulations regarding obtaining medical licensing in Botswana, we were able to screen 89 patients and of those recruit 76 (86%) for our study. For the purpose of power calculations, our original study design anticipated each patient having a single mucocutaneous problem. However, we found that multiple patients presented with multiple mucocutaneous conditions, with a mean of 2.1 diagnoses per enrolled patient, yielding a total of 159 diagnoses. (Table 1, Figure 1) We decided therefore to proceed with our analyses and reporting of the data, as described below, with each diagnosis analyzed as a separate photocase. Median age was 39 years (IQR 32-45). Fifty-six percent (n=43) were females.

Descriptive data

At the time of the study, Evaluator 1, i.e., the Face to Face (FTF) evaluator, had been practicing as a board-certified dermatologist for one year; Evaluator 2 was board-certified in dermatology for 6 months; Evaluator 3 was board-certified in oral medicine/dentistry for 41years and has expertise in oral lesions in HIV-positive patients; Evaluator 4 was board-certified in dermatology for 4 years; and Evaluator 5 was board-certified in dermatology for 2 years. In terms of clinical time in Botswana, Evaluators 1 and 4 had spent multiple several week periods performing clinical evaluations in Botswana, while Evaluators 2 and 5 had each experienced 1 clinical trip to Botswana lasting several weeks.

Outcome data

Table 1 describes the summary numbers of diagnoses and diagnostic categories found by each reviewer per photocase. The majority of photos were thought to contain only one diagnosis; however a number of photos were thought by each of the reviewers to represent more than one diagnosis. While the face to face interviewer reported 159 diagnoses among the 76 enrolled patients, the remote evaluators all found varying numbers of diagnoses when looking at the same photos (154-313 for the teledermatology mobile evaluators). Among the 28 cases reported by the face-to-face evaluator to have oral pathology, 39 diagnoses were reported by Reviewer 3. Furthermore, for each diagnosis, the reviewer also had an opportunity to provide a differential diagnosis. Each of the evaluators had a different approach with some evaluators (i.e., the face-to-face evaluator, Reviewer 2, Reviewer 4) providing extensive lists of possible differential diagnoses for a number of cases.

Main Results

Table 2 describes the test-retest reliability of our main outcomes. Reviewers agreed with their own previous primary diagnoses 52-80% of the time (kappa Reviewer 2: 0.47, 95% CI 0.35-0.59, kappa Reviewer 5: 0.78, 95% CI 0.67-0.88). Agreement on diagnostic category for the primary diagnosis varied over time from 36% for Reviewer 2 (kappa 0.29, 95% CI 0.18-0.42) to 77% for Reviewer 5 (kappa 0.73, 95% CI 0.61-0.84) Test-retest agreement for management choices ranged from 55% for Reviewer 2 (kappa 0.17, 95% CI -0.01-0.36) to 69% for Reviewer 5 (kappa 0.54, 95% CI 0.38-0.70).

Table 3 describes the inter-rater reliability of our main outcomes. Agreement between the face to face evaluator and the remote evaluators for the primary diagnosis ranged from 47% for Reviewer 2 (kappa 0.41, 95% CI 0.31-0.52) to 57% for Reviewer 4 (kappa 0.51, 95% CI 0.41-0.61). Agreement on the diagnostic category to which the primary diagnosis belonged ranged from 29% for Reviewer 2 (kappa 0.22, 95% CI 0.14-0.31) to 50% for Reviewer 5 (kappa 0.43, 95% CI 0.34-0.53). Agreement between the face to face interviewer and the remote evaluators on how to manage the patient's primary diagnosis ranged from 32% for Reviewer 2 (0.08, 95% CI 0.02-0.15), to 51% for Reviewer 4 (kappa 0.12, 95% CI 0.01-0.23). When looking only at the subset of cases with oral lesions, inter-rater agreement ranged from 62 to 68% for the primary differential diagnosis. Kappa coefficients ranged from 0.51 to 0.58 for diagnosis, 0.17 to 0.55 for diagnostic category and -0.14 to 0.09 for management. The ten primary diagnoses most often made by the face-to-face evaluator were: other (N=34), Kaposi sarcoma (N=20), HSV (N=10), acne (N=8), condyloma (N=6), atopic dermatitis (N=5), candidiasis (N=5), superficial fungal infection (not including nails (N=5), verruca vulgaris (N=4), and xerosis (N=4). Sensitivity and Specificity for these conditions are reported in Table 4.

Discussion

Our protocol for mobile phone mediated store and forward evaluations resulted in varying diagnostic conclusions among different evaluators (inter-rater variability) and over time for the same evaluator (intra-rater variability). Kappa values have previously been described by expert opinion as less than 0 indicating no agreement, between 0-.20 indicating slight

agreement (i.e. poor agreement), .21-.40 as fair agreement, .41-.60 as moderate agreement, .61-.80 as substantial agreement (i.e., good agreement), and .81-1 as almost perfect agreement.[12] To provide context within dermatology, inter-observer agreement for the histologic diagnosis of melanoma has been reported to vary in the fair to good range in several studies.[13-17] In our study test-retest reliability for the diagnosis was moderate to good (kappa coefficients ranging from 0.47 to 0.78); fair to good for diagnostic category (kappa range 0.29 to 0.73), and poor to moderate for management (kappa range 0.17 to 0.54). Furthermore, while it appears that diagnosis and diagnostic category, achieve at the least, fairly reliable or better inter-rater responses, management choices made by the majority of individual evaluators cannot be relied upon for consistency over time. Of note, Reviewers 4 and 5 consistently achieved the highest levels of test-retest reliability among all categories of measurement; these reviewers were the ones with the most clinical (both face-to-face and teledermatology-based) exposure to this specific population of patients. Furthermore, the next most consistent evaluator, Reviewer 3 also had significant years of clinical experience in dealing with HIV positive patients, although this experience was based largely in an urban tertiary care setting in Philadelphia, not in a sub-Saharan population. Taken together, these observations suggest that further research is warranted into determining the optimal amount of clinical experience that each tele-evaluator should have in working within their target patient population to achieve consistent remote diagnostic evaluations.

When looking at inter-rater reliability of our main outcomes, agreement was highest across all reviewers for diagnosis (47-57%) with kappas consistently in the moderate range (0.41-0.51), and even higher and more consistent when the cases were limited to those with oral pathology (62-68% agreement, kappa range 0.51-.58). Thus, it appears that both dermatologists and oral medicine specialists appear to be able to diagnose the oral lesions encountered within this population with the same level of consistency. With the exception of Reviewer 2, inter-rater reliability seemed in the fair to moderate range for diagnostic category of the primary diagnosis overall and in the oral case subset. Diagnostic accuracy has been noted to be correlated with degree of clinical experience.[18] As noted earlier, Reviewer 2 had the least amount of teledermatology experience among all reviewers at the time of this study. It is possible therefore, that a combination of further in-person clinical or mobile teledermatology exposure in this setting could yield more accurate results.

Management suggestions made by the mobile evaluators were very poorly correlated with the management choices of the face to face dermatologist. Mobile evaluators more frequently recommended management options that included further diagnostic testing (i.e., test or test and treat), whereas the face-to-face evaluator more frequently chose management options involving treatment alone due to technical limitations in the availability of diagnostic equipment (e.g., biopsy kits) and the duration of time it took to obtain definitive results in this clinical setting.

Due to the imperfect nature of using a single face-to-face evaluation as our gold standard, we also looked at inter-rater agreement among the mobile evaluators themselves. For the primary differential diagnosis, the mobile evaluators achieved moderate reliability for diagnosis with a kappa statistic of 0.44 (95% CI 0.36 - 0.52), which increased to good

reliability, Kappa =0.61 (95% 0.48 - 0.75), when limited to the cases with oral involvement. For diagnostic category of the primary differential diagnosis, the inter-rater reliability of the remote evaluators among themselves reached Kappa= 0.34 (95% CI 0.27 - 0.42) overall and kappa =0.37 (95% CI 0.25 - 0.52) for the oral cases. Finally, agreement on management of the primary diagnosis was poor among the remote evaluators, with kappa=0.04 (95% CI -0.04 - 0.12) for the dermatology based evaluators, and kappa =0.01 (95% CI -0.13 - 0.18) among the oral evaluations. Broadly speaking, these levels of agreement are similar to the level of agreement each mobile evaluator achieved when compared to the face-to-face evaluation.

In terms of validity, with the exception of Reviewer 3's evaluation of herpes simplex, specificity for the primary diagnosis outstripped sensitivity for the top ten most common conditions identified by the face-to-face evaluator. These findings suggest that mobile teleconsultants may be better at ruling in the ten most common diagnoses in this population than they are at ruling them out (i.e., higher specificity than sensitivity). The utility of mobile teledermatology may vary depending on the mucocutaneous conditions for which the consultation is sought, however. For example, in a study of Spanish patients attending a pigmented lesion clinic, the sensitivity for the detection of malignant versus benign tumors by teledermatology was 0.99 (95% CI, 0.98-1.00) while specificity was 0.62 (95% CI, 0.56-0.69).[19] Further research into which conditions are better suited for teledermatology consultation, and mobile teledermatology in particular, is warranted.

In this pilot study, due to logistical limitations, the nurse interviewed and photographed patients after the face-to-face evaluation was completed. As the nurse also doubled as the translator in our study, and was at times present for the face-to-face evaluation, this approach may have led to photographs that were more likely to contain key recognizable and diagnostic aspects of a mucocutaneous condition. To limit this bias, future studies may consider timing the nurse interview and photography prior to the face-to-face evaluations.

Some of the differences we observed in inter-rater reliability of the main outcomes may be accounted for by the difficulty of defining certain conditions into one specific category. For example, while most of the mobile evaluators easily identified Kaposi sarcoma as the likely diagnosis in a given photocase, they often differed when defining its diagnostic category, with some consistently choosing neoplasm while others categorizing it as a viral infection. Our intention in including diagnostic category as an outcome in this study stemmed from the rationale that in the real world setting, should a definite diagnosis not be apparent through mobile teledermatology consultation, suggesting a diagnostic category might at least allow triage of the patient in terms of broad management steps (i.e biopsy, refer for face-to-face dermatology evaluation or empiric treatment). Although the diagnostic categories that we listed have face validity, (e.g., were drawn from a gold-standard textbook of dermatology[20]) and had good intra-reliability and fair to moderate reliability among the more experienced evaluators, it is evident from our findings that more work needs to be done in terms of establishing construct validity to demonstrate the utility of these categories. We hope nonetheless that detailing our findings here will contribute to the design of evaluation measures in future studies. While the face to face evaluator had immediate and direct access to patient clinical information, i.e., history and examination, the remote evaluations were

limited by at least two factors: a mobile phone interface that limited history intake to demographic, yes/no or multiple choice based questions and a data collector with very limited knowledge of dermatology and dermatologic photography. For instance, when asked to rate the quality of the photos they received as either "good"=1, "satisfactory"=2, or "poor"=3, the mean photographic quality rating for most of the mobile evaluators fell between satisfactory and poor (Mean test/retest values: Rater 2= 2.1/2.3, Rater 3= 2.5/2.6, Rater 4=1.4/1.4, Rater 5= 2.4/2.4).

It is likely that with more thorough access to history and better quality of photographs, the sensitivity of the diagnoses will increase. With this in mind, future versions of the mobile teledermatology phone software will include the ability to incorporate free text, which will allow this hypothesis to be further tested. However, the data collection limitations imposed by a data collection intermediary with limited dermatology knowledge will likely remain; most health care workers seeking mobile teledermatology consultations in the real world setting are not specifically trained in taking or equipped to take photographs in a way that is ideal for mucocutaneous evaluation. For example, even with phones equipped with high quality cameras, photos may be taken during rounds in the wards or in clinics where ambient lighting is less than ideal for skin and especially for mucosal or intraoral photography. It is possible that due to the training given to our nurse, our results are biased towards more accurate diagnosis than what would be expected from an individual completely untrained in dermatologic photography. However, we anticipate that more extensive dermatologic and photographic training of the health care workers who use the software to upload history and photos in the real world setting may alleviate some of these limitations by improving the quality of the data that is captured.

In conclusion, as evidenced by our pilot study, although the introduction of mobile teledermatology into resource limited population of HIV positive patients in sub-Saharan Africa has significant theoretical potential for improving access to care, much work is still needed to optimize and validate the use of this technology on a larger scale in this population. While others have addressed the diagnostic accuracy and clinical outcomes of both store and forward as well as live video-based teledermatology, to our knowledge, ours is the first attempt at validating mobile teledermatology in this practice setting. Several questions arise as a result of our findings. For example, what kind of and how much training is necessary to perform mobile teledermatology evaluations? How can the mobile technology be used to maximize the validity of the mobile consultation? What kind of training is required to adequately capture and transmit the data? Is receiving diagnostic advice from a remote source sufficient guidance if management advice is unreliable or not technically feasible? Future studies will need to address these concerns. Finally, a cost-benefit analysis of this work in this clinical setting or on a larger scale may need to be determined before this promising advance in technology can be used to fill the gap between the need for dermatologic care and the number of qualified and available providers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Schmid-Grendelmeir P, Doe P, Pakenham-Walsh N. Teledermatology in sub-Saharan Africa. Curr Probl Dermatol. 2003; 32:233–246. [PubMed: 12472018]

2. Morrone A. Poverty, health and development in dermatology. Int J Dermatol. 2007; 46(Suppl 2):1–9. [PubMed: 17958622]

3. Spira R, Mignard M, Doutre M, Morlat P, Dabis F. Prevalence of cutaneous disorders in a population of HIV-infected patients. Arch Dermatol. 1998; 134:1208–1212. [PubMed: 9801675]

4. Berger TG, Obuch ML, Goldschmidt RH. Dermatologic manifestations of HIV infection. Am Fam Physician. 1990 Jun; 41(6):1729–42. [PubMed: 2190455]

5. Goh BK, Chan RK, Sen P, et al. Spectrum of skin disorders in human immunodeficiency virus-infected patients in Singapore and the relationship to CD4 lymphocyte counts. Int J Dermatol. 2007; 46:695–9. [PubMed: 17614796]

6. Ebner C, Wurm EM, Binder B, et al. Mobile teledermatology: a feasibility study of 58 subjects using mobile phones. J Telemed Telecare. 2008; 14:2–7. [PubMed: 18318921]

7. Chung P, Yu T, Scheinfeld N. Using cellphones for teledermatology, a preliminary study. Dermatology Online Journal. 2007; 13

8. Tran K, Ayad M, Weinberg J, Cherng A, Chowdhury M, Monir S, El Hariri M, Kovarik C. Mobile teledermatology in the developing world: implications of a feasibility study on 30 Egyptian patients with common skin diseases. J Am Acad Dermatol. 2011 Feb; 64(2):302–9. [PubMed: 21094560]

9. Massone C, Lozzi GP, Wurm E, et al. Cellular phones in clinical teledermatology. Arch Dermatol. 2005; 141:1319–20. [PubMed: 16230577]

10. Braun RP, Vecchietti JL, Thomas L, et al. Telemedical wound care using new generation of mobile telephones: a feasibility study. Arch Dermatol. 2005; 141:254–8. [PubMed: 15724023]

11. Azfar RS, Weinberg JL, Cavric G, Lee-Keltner IA, Bilker WB, Gelfand JM, Kovarik CL. HIV-positive patients in Botswana state that mobile teledermatology is an acceptable method for receiving dermatology care. J Telemed Telecare. 2011 Aug 15. Epub ahead of print.

12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar; 33(1):159–74. [PubMed: 843571]

13. Braun RP, Gukowicz-Krusin D, Rabinovitz H, et al. Agreement of dermatopathologists in the evaluation of clinically difficult melanocytic lesions: how golden is the 'gold standard'? Dermatol. 2012; 224:51–8.

14. Farmer ER, Gonin R, Hanna M. Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists. Hum Pathol. 1996 Jun; 27(6):528–31. [PubMed: 8666360]

15. Corona R, Mele A, Amini M, et al. Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. J Clin Oncol. 1996; 14:1218–1223. [PubMed: 8648377]

16. CRC Melanoma Pathology Panel. A nationwide survey of observer variation in the diagnosis of thin cutaneous malignant melanoma including the MIN terminology. J Clin Pathol. 1997; 50:202–205. [PubMed: 9155669]

17. Lodha S, Saggar S, Celebi JT, et al. Discordance in the histopathologic diagnosis of difficult melanocytic neoplasms in the clinical setting. J Cutan Pathol. 2008; 35:349–352. [PubMed: 18333894]

18. Levin YS, Warshaw EM. Teledermatology: A Review of Reliability and Accuracy of Diagnosis and Management. Dermatologic Clinics. 2009 Apr; 27(2):163–76. [PubMed: 19254660]

19. Moreno-Ramirez D, Ferrandiz L, Nieto-Garcia A, et al. Store-and-forward teledermatology in skin cancer triage: experience and evaluation of 2009 teleconsultations. Arch Dermatol. 2007; 143(4):479–84. [PubMed: 17438180]

20. James, WD.; Berger, Berger T.; Elston, DE. Andrews' Diseases of the Skin: Clinical Dermatology. 10th. Philadelphia: WB Saunders; 2006.

89 Patients Assessed for Eligibility

13 Excluded:
4 Died or left before consent signed
4 Did not meet study criteria
5 Refused to participate

76 Subjects Enrolled

1 Subject contributed 4 photocases

52 Subjects contributed 1 photocase

14 Subjects contributed 2 photocases

9 Subjects contributed 3 photocases
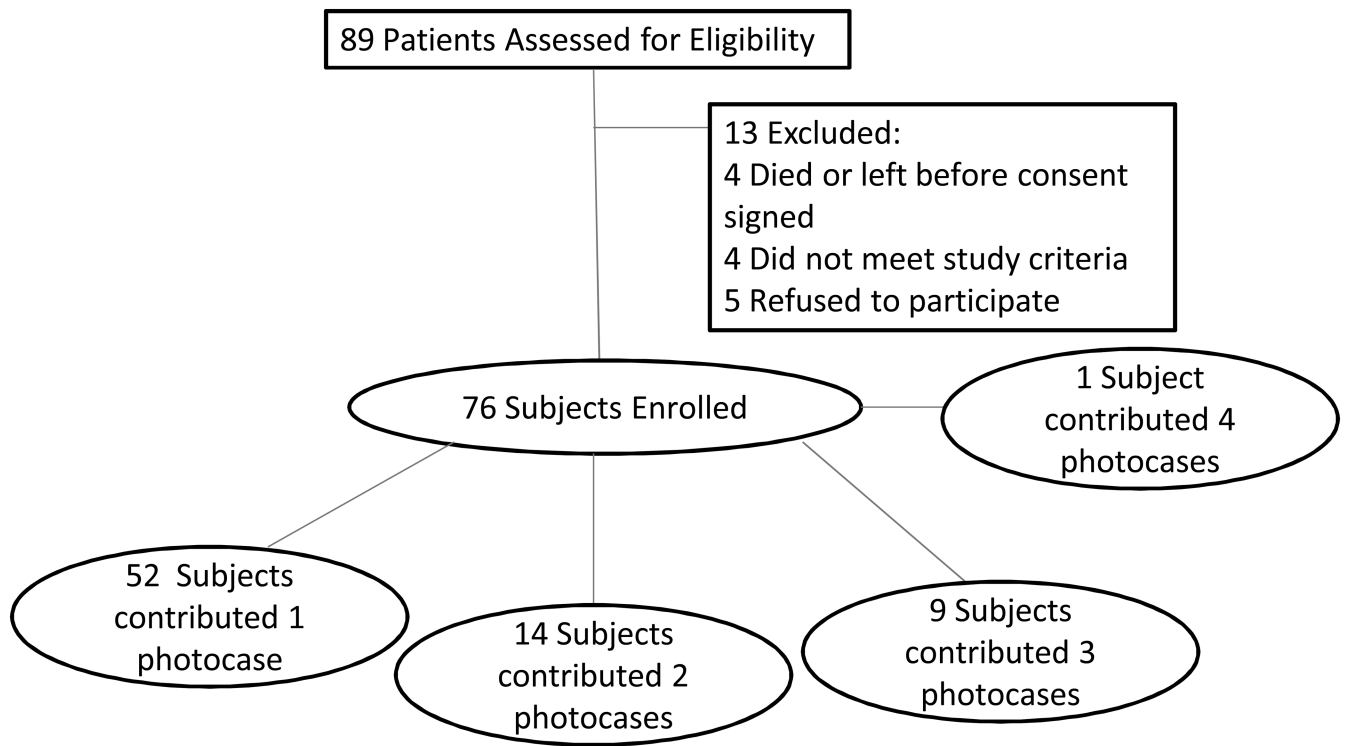
Figure 1. Subject and photocase allocation

NIH-PA Author Manuscript   NIH-PA Author Manuscript   NIH-PA Author Manuscript

Table 1

Photocase characteristics*

	FTF reviewer (Gold Standard)	Reviewer 2	Reviewer 4	Reviewer 5	Reviewer 3 (oral cases only)
Total # of diagnoses made by each evaluator *	159	239	313	154	39
# of photocases with 1 diagnosis	97	154	195	105	24
# of photocases with at least 2 diagnoses	37	57	83	35	10
# of photocases with at least 3 diagnoses	18	21	32	11	4
# of photocases with at least 4 diagnoses	5	7	3	3	1
# of photocases with no differential diagnosis	126	114	121	111	33
# of photocases with at least two differential diagnoses	32	65	92	37	5
# of photocases with at least three differential diagnoses	14	38	61	6	1
# of photocases with at least four differential diagnoses	3	14	28	-	-
# of photocases with at least five differential diagnoses	1	6	7	-	-
# of photocases with at least six differential diagnoses	1	2	3	-	-

* Each patient could have multiple conditions during the face-to-face evaluation. Each of these conditions was considered separately and sent to the mobile evaluators as a separate photocase. However, when each photocase was evaluated by the remote physician, it was possible that the remote physician may consider that photocase to contain multiple diagnoses (summarized in the top half of the table) Furthermore, evaluators could provide a differential diagnosis for any condition that they determined was present (summarized in the bottom half of the table).

Table 2

Test-Retest Reliability of Main Outcomes

	Reviewer 2	Reviewer 4	Reviewer 5	Reviewer 3 (oral cases only)
% Test-retest Agreement by "diagnosis" for the primary differential diagnosis (N)	52.00 (40)	76.83 (63)	80.00 (59)	61.11 (14)
% Test-retest Agreement by "diagnostic category for the primary differential diagnosis (N)	36.49 (28)	74.39 (61)	76.81 (56)	66.67 (14)
% Test-retest Agreement by "management" (N)	54.93 (93)	67.95 (136)	69.44 (63)	57.14 (24)
Kappa "diagnosis" for the primary differential diagnosis (95% CI)	0.47 (0.35-0.59)	0.73 (0.63-0.83)	0.78 (0.67-0.88)	0.47 (0.22-0.79)
Kappa "diagnostic category" for the primary differential diagnosis (95% CI)	0.29 (0.18-0.42)	0.70 (0.59-0.81)	0.73 (0.61-0.84)	0.52 (0.26-0.82)
Kappa "management" (95% CI)	0.17 (-0.01-0.36)	0.48 (0.32-0.64)	0.54 (0.38-0.70)	0.32 (-0.07-0.75)

Table 3
Inter-rater Reliability of Main Outcomes

	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
Inter-rater Agreement by "diagnosis" for the primary differential diagnosis (%) (N=136)	46.6		56.8	48.6
Inter-rater Agreement by "diagnostic category" for the primary differential diagnosis (%) (N=136)	28.7		43.2	50.0
Inter-rater Agreement by "managment" (%) (N=234)	31.6		50.8	40.2
Kappa "diagnosis" for the primary differential diagnosis (95% CI)	0.41 (0.31-0.52)		0.51 (0.41-0.61)	0.43 (0.34-0.53)
Kappa "diagnostic category" for the primary differential diagnosis (95% CI)	0.22 (0.14-0.31)		0.37 (0.27-0.46)	0.43 (0.34-0.53)
Kappa "management" (95% CI)	0.08 (0.02-0.15)		0.12 (0.01-0.23)	0.04 (-0.06- 0.15)
		Oral Cases Only:		
	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
Inter-rater Agreement by "diagnosis" for the primary differential diagnosis (%) (N=29)	66.67	68.0	66.67	61.54
Inter-rater Agreement by "diagnostic category" for the primary differential diagnosis (%) (N=29)	34.62	64.0	66.67	64.00
Inter-rater Agreement by "managment" (%) (N=30)	32.41	44.4	46.15	34.62
Kappa "diagnosis" for the primary differential diagnosis (95% CI)	0.58 (0.39-0.79)	0.58 (0.36-0.80)	0.57 (0.36-0.78)	0.51 (0.33-0.73)
Kappa "diagnostic category" for the primary differential diagnosis (95% CI)	0.17 (0.041-0.36)	0.51 (0.32-0.74)	0.55 (0.37-0.76)	0.52 (0.32-0.73)
Kappa "management" (95% CI)	-0.010 (-0.21-0.13)	0.09 (-0.19-0.41)	-0.03 (-0.28-0.22)	-0.14 (-0.38- 0.09)

Table 4

Validity: Sensitivity and Specificity for Primary Diagnosis

Diagnosis	Reviewer 2		Reviewer 3		Reviewer 4		Reviewer 5	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Acne (n=8)	0.5(0.40-0.60)	1(0.96-1)			0.625(0.60-0.65)	1(0.96-1)	0.38(0.35-0.41)	1(0.96-1)
Atopic dermatitis (n=5)	0.4(0.31-0.50)	0.95(0.94-0.97)			0.2(0.13-0.29)	0.96(0.94-0.97)	0(0-0.04)	1(0.96-1)
Candidiasis (n=5)	0.2(0.13-0.29)	0.99(0.98-0.99)	0.25(0.18-0.34)	0.84(0.81-0.86)	0.2(0.13-0.29)	0.98(0.97-0.99)	0.4(0.31-0.50)	0.96(0.90-0.98)
Condyloma accuminata (n=6)	0.17(0.15-0.19)	0.95(0.93-0.96)			0.83(0.81-0.85)	1(0.96-1)	0.83(0.81-0.85)	1(0.96-1)
Xerosis (n=4)	0(0-0.04)	0.98(0.97-0.98)			0(0-0.04)	0.99(0.99-1.0)	0(0-0.04)	0.98(0.97-0.99)
Dermatophytosis (not nail) (n=5)	0.6(0.50-0.69)	0.98(0.97-0.98)			0.2(0.13-0.29)	0.97(0.96-0.98)	0.8(0.71-0.87)	0.95(0.94-0.96)
Herpes simplex (n=10)	0.7(0.60-0.78)	0.97(0.96-0.98)	0.88(0.85-0.89)	0.85(0.83-0.87)	0.8(0.71-0.87)	0.98(0.97-0.99)	0.7(0.60-0.78)	0.96(0.94-0.97)
Kaposi's sarcoma (n=20)	0.6(0.50-0.69)	0.98(0.97-0.99)	0.78(0.75-0.80)	1(0.96-1)	0.85(0.77-0.91)	0.95(0.94-0.96)	0.8(0.71-0.87)	0.98(0.97-0.99)
Other (n=34)	0.29(0.27-0.32)	0.86(0.84-0.88)	0.2(0.13-0.29)	0.93(0.92-0.95)	0.53(0.50-0.56)	0.86(0.84-0.88)	0.27(0.24-0.29)	0.89(0.86-0.90)
Verruca vulgaris (n=4)	0.25(0.18-0.34)	0.99(0.98-0.99)			0.5(0.40-0.60)	0.99(0.99-1.0)	0.5(0.40-0.60)	1(0.96-1)