Reliability and Validity of Mobile Teledermatology in Human Immunodeficiency Virus–Positive Patients in Botswana: A Pilot Study

Rahat S. Azfar, MD, MSCE; Robert A. Lee, MD, PhD; Leslie Castelo-Soccio, MD, PhD; Martin S. Greenberg, DDS; Warren B. Bilker, PhD; Joel M. Gelfand, MD, MSCE; Carrie L. Kovarik, MD

In many parts of the world, particularly in sub-Saharan Africa, there is a severe shortage of dermatologic specialists. Dermatologic care is often provided by clinicians and rural health workers with limited training in dermatologic care. In these regions, this shortage is felt more acutely in communities with high rates of human immunodeficiency virus (HIV), since there is an increased burden of both prevalence and severity of skin and mucosal disease in this group compared with the immunocompetent population. In addition, the presence of several particular mucocutaneous conditions may also affect HIV management.

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 uses cellular phone networks, which are more stable and accessible, to perform store-and-forward teledermatology consultations. While several studies have evaluated diagnostic agreement, relatively few have investigated the reliability and validity of mobile teledermatology compared with the gold standard of face-to-face evaluation by a dermatologist. Moreover, to our knowledge, this technology has not been tested in sub-Saharan Africa among HIV-positive patients.

We sought to determine whether the use of mobile teledermatology technology in HIV-positive patients in Gaborone, Botswana, was reliable and produced valid assessments compared with face-to-face dermatologic consultations. We hypothesized that health care workers could
transmit clinical information and photographs through a cellular phone, which would allow reliable and valid remote dermatologic evaluations similar in quality to in-person consultations.

**Methods**

**Study Design and Setting**

We conducted a cross-sectional pilot study of adult patients in Botswana with HIV and mucocutaneous disease. The study was approved by the institutional review boards at the University of Pennsylvania, Princess Marina Hospital, and the Botswana Ministry of Health.

**Participants**

The study was conducted in consecutively recruited HIV-positive patients at least 18 years old with a skin or mucosal condition that had not been evaluated by a dermatologist. The patients were recruited from the medical and oncologic wards, the dermatologic clinic, and the infectious disease clinic at Princess Marina Hospital in Gaborone, Botswana; from the 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, from August 20 through September 21, 2009.

**Data Sources/Measurement and Bias**

All patients received a face-to-face clinical evaluation by a US-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 cellular phone encounter. A Setswana-speaking nurse obtained oral and written consent and clarified any patient questions. Enrolled patients received P (pula) 30 (US $4) to cover the cost of their travel at the end of the mobile encounter. The face-to-face dermatologist completed a separate deidentified clinical evaluation form to collect study data for each enrolled patient (eFigure 1 in the Supplement). This evaluation was used as the gold standard for comparative purposes. Patients who consented to the cellular phone encounter were then seen by the nurse interviewer, who collected their data and forwarded them for the mobile teledermatology evaluation. To simulate a typical hypothetical setting in which mobile teledermatology may be used, the nurse, who had no previous experience in dermatology, was trained in using the cellular phone software for history taking and medical photography a few days before beginning the study but worked independently from the face-to-face dermatologist to collect data. Data were gathered without any personal identifying information and forwarded directly from the Samsung Soul SGH-U900 cellular phone with a 5-megapixel camera to a secure password-protected teledermatology evaluation website (eFigures 2, 3, and 4 in the Supplement).

After initial data collection was completed, mobile evaluations were completed by 3 US-based board-certified dermatologists (L.C.-S., C.L.K, and R.A.L.) and 1 board-certified oral medicine specialist (M.S.G.) (eFigure 5 in the Supplement). Each evaluator had varied levels of clinical experience working in the sub-Saharan HIV-positive population or similar populations. The oral medicine specialist assessed only those cases that included oral pathologic conditions. Mobile evaluations were not used to guide clinical decision making and were conducted solely 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 (bacterial infection, neoplasm, papulosquamous inflammation, etc), and their recommendations for management (treat, test, test and treat, or refer for face-to-face evaluation). Test-retest reliability was assessed several months after the initial mobile evaluations were completed by giving the mobile evaluators 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 United States in the months preceding the beginning of the study.

**Variables and Statistical Analysis**

We calculated descriptive statistics for the overall cohort as well as interrater and test-retest reliability for each main outcome and sensitivity and specificity for each diagnosis using Stata 10.1 (StataCorp LP). For the reliability analyses, our main outcomes were the κ 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 interrater reliability and validity of the diagnoses. The significance level (α) 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 condition.

**Results**

**Participants**

Patient characteristics have been described previously. Due to the loss of more than 1 week of allocated study time to recently placed government regulations regarding obtaining medical licensing in Botswana, we were able to screen 89 patients, of whom we recruited 76 (85%) for our study. For the purpose of power calculations, our original study design anticipated each patient having a single mucocutaneous condition. However, a number of patients had multiple mucocutaneous conditions, with a mean of 2.1 diagnoses per enrolled patient, yielding 159 diagnoses (Table 1 and Figure). 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 (interquartile range, 32-45 years). Forty-three (57%) were women.
Mobile Teledermatology in HIV-Positive Patients

Original Investigation Research

Mary diagnosis varied over time from 36.5% for evaluator 2 (κ, 0.35 to 0.59) from evaluator 2 and 0.78 (0.67 to 0.88) from evaluator 2 (κ, 0.41; 95% CI, 0.31 to 0.52) to 57% for evaluator 4 (0.51; 0.41 to 0.61). Agreement on diagnostic category for the primary diagnosis varied from 29% for evaluator 2 (κ, 0.17; 95% CI, –0.01 to 0.36) to 69% for evaluator 5 (0.54; 0.38 to 0.70).

Table 3 describes the interrater reliability of our main outcomes. Agreement between the face-to-face dermatologist and the remote reviewers for the primary diagnosis ranged from 47% for evaluator 2 (κ, 0.41; 95% CI, 0.31 to 0.52) to 57% for evaluator 4 (0.51; 0.41 to 0.61). Agreement on the diagnostic category to which the primary diagnosis belonged ranged from 29% for evaluator 2 (κ, 0.22; 95% CI, 0.14 to 0.31) to 50% for evaluator 5 (0.43; 0.34 to 0.53). Agreement between the face-to-face dermatologist and the remote reviewers on how to treat the patient’s primary diagnosis ranged from 32% for evaluator 2 (κ, 0.08; 95% CI, 0.02 to 0.15) to 51% for evaluator 4 (0.12; 0.01 to 0.23). When looking at only the subset of cases with oral lesions, interrater agreement ranged from 62% to 68% for the primary differential diagnosis. The x 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 10 primary diagnoses made most often by the face-to-face dermatologist were other (n = 34), Kaposi sarcoma (n = 20), herpes simplex virus (n = 10), acne (n = 8), condyloma acuminate (n = 6), atopic dermatitis (n = 5), candidiasis (n = 5), dermatophytosis (not including nails, n = 5), verruca vulgaris (n = 4), and xerosis (n = 4). Sensitivity and specificity for these conditions are reported in Table 4.

Table 1. Photocase Characteristics

Table 2 describes the test-retest reliability of our main outcomes. Reviewers agreed with their own previous primary diagnoses 52% to 80% of the time, with x (95% CI) values of 0.47 (0.35 to 0.59) from evaluator 2 and 0.78 (0.67 to 0.88) from evaluator 5. Agreement on diagnostic category for the primary diagnosis varied over time from 36.5% for evaluator 2 (κ, 0.29; 95% CI, 0.18 to 0.42) to 77% for evaluator 5 (0.73; 0.61 to 0.84). Test-retest agreement for management choices ranged from 55% for evaluator 2 (κ, 0.17; 95% CI, –0.01 to 0.36) to 69% for evaluator 5 (0.54; 0.38 to 0.70).

Table 4

The x 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 10 primary diagnoses made most often by the face-to-face dermatologist were other (n = 34), Kaposi sarcoma (n = 20), herpes simplex virus (n = 10), acne (n = 8), condyloma acuminate (n = 6), atopic dermatitis (n = 5), candidiasis (n = 5), dermatophytosis (not including nails, n = 5), verruca vulgaris (n = 4), and xerosis (n = 4). Sensitivity and specificity for these conditions are reported in Table 4.

Descriptive Data

At the time of the study, evaluator 1 (ie, the face-to-face evaluator) had been practicing as a board-certified dermatologist for 1 year; evaluator 2 had been board certified in dermatology for 6 months; evaluator 3 had been board certified in oral medicine and dentistry for 41 years, with expertise in oral lesions in HIV-positive patients; evaluator 4 had been board certified in dermatology for 4 years; and evaluator 5 had been 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 assessments 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. Most photographs were thought to contain only 1 diagnosis; however, each evaluator believed that some represented more than 1 diagnosis. While the face-to-face dermatologist reported 159 diagnoses among the 76 enrolled patients, the remote evaluators all found varying numbers of diagnoses when examining the same photographs (154-313 for the teledermatology mobile evaluators). Among the 28 cases reported by the face-to-face dermatologist to have an oral pathologic condition, evaluator 3 reported 39 diagnoses. Furthermore, evaluators could provide a differential diagnosis. Each evaluator had a different approach, with some (ie, evaluators 1, 2, and 4) providing extensive lists of possible differential diagnoses for each case.

Main Results

Table 2 describes the test-retest reliability of our main outcomes. Reviewers agreed with their own previous primary diagnoses 52% to 80% of the time, with x (95% CI) values of 0.47 (0.35 to 0.59) from evaluator 2 and 0.78 (0.67 to 0.88) from evaluator 5. Agreement on diagnostic category for the primary diagnosis varied over time from 36.5% for evaluator 2 (κ, 0.29; 95% CI, 0.18 to 0.42) to 77% for evaluator 5 (0.73; 0.61 to 0.84). Test-retest agreement for management choices ranged from 55% for evaluator 2 (κ, 0.17; 95% CI, –0.01 to 0.36) to 69% for evaluator 5 (0.54; 0.38 to 0.70).

Table 3 describes the interrater reliability of our main outcomes. Agreement between the face-to-face dermatologist and the remote reviewers for the primary diagnosis ranged from 47% for evaluator 2 (κ, 0.41; 95% CI, 0.31 to 0.52) to 57% for evaluator 4 (0.51; 0.41 to 0.61). Agreement on the diagnostic category to which the primary diagnosis belonged ranged from 29% for evaluator 2 (κ, 0.22; 95% CI, 0.14 to 0.31) to 50% for evaluator 5 (0.43; 0.34 to 0.53). Agreement between the face-to-face dermatologist and the remote reviewers on how to treat the patient’s primary diagnosis ranged from 32% for evaluator 2 (κ, 0.08; 95% CI, 0.02 to 0.15) to 51% for evaluator 4 (0.12; 0.01 to 0.23). When looking at only the subset of cases with oral lesions, interrater agreement ranged from 62% to 68% for the primary differential diagnosis. The x 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 10 primary diagnoses made most often by the face-to-face dermatologist were other (n = 34), Kaposi sarcoma (n = 20), herpes simplex virus (n = 10), acne (n = 8), condyloma acuminate (n = 6), atopic dermatitis (n = 5), candidiasis (n = 5), dermatophytosis (not including nails, n = 5), verruca vulgaris (n = 4), and xerosis (n = 4). Sensitivity and specificity for these conditions are reported in Table 4.

Table 1. Photocase Characteristics

Table 2. Main Results

Table 4

Table 4

Abbreviation: FTF, face-to-face.
Discussion

Our protocol for cellular phone–mediated store-and-forward evaluations resulted in varying diagnostic conclusions among different evaluators (interrater variability) and over time for the same evaluator (intrarater variability). The κ coefficients have been described by expert opinion as less than 0, indicating no agreement; 0 to 0.20, slight agreement (ie, poor agreement); 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement (ie, good agreement); and 0.81 to 1.00, almost perfect agreement.12 To provide context within dermatology, interobserver agreement for the histologic diagnosis of melanoma has been reported to vary from fair to good in several studies.13-17 In our study, test-retest reliability was moderate to good for the diagnosis (κ range, 0.47–0.78), fair to good for diagnostic category (0.29–0.73), and poor to moderate for management (0.17–0.54). Furthermore, while it appears that diagnosis and diagnostic category achieve fairly reliable or better interrater responses, management choices made by most individual evaluators cannot be relied on for consistency over time. Of note, evaluators 4 and 5 consistently achieved the highest levels of test-retest reliability among all categories of measurement; these reviewers had the most clinical (both face-to-face and teledermatology-based) exposure to this specific population of patients. Furthermore, the next most consistent evaluator, evaluator 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

Table 2. Test-Retest Reliability of Main Outcomes

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Evaluator 2</th>
<th>Evaluator 4</th>
<th>Evaluator 5</th>
<th>Evaluator 3 (Oral Cases Only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest agreement by diagnosis for the primary differential diagnosis, No. (%)</td>
<td>40 (52.0)</td>
<td>63 (76.8)</td>
<td>59 (80.0)</td>
<td>14 (61.1)</td>
</tr>
<tr>
<td>Test-retest agreement by diagnostic category for the primary differential diagnosis, No. (%)</td>
<td>28 (36.5)</td>
<td>61 (74.4)</td>
<td>56 (76.8)</td>
<td>14 (66.7)</td>
</tr>
<tr>
<td>Test-retest agreement by management, No. (%)</td>
<td>93 (54.9)</td>
<td>136 (68.0)</td>
<td>63 (69.4)</td>
<td>24 (57.1)</td>
</tr>
<tr>
<td>Diagnosis for the primary differential diagnosis, κ (95% CI)</td>
<td>0.47 (0.35 to 0.59)</td>
<td>0.73 (0.63 to 0.83)</td>
<td>0.78 (0.67 to 0.88)</td>
<td>0.47 (0.22 to 0.79)</td>
</tr>
<tr>
<td>Diagnostic category for the primary differential diagnosis, κ (95% CI)</td>
<td>0.29 (0.18 to 0.42)</td>
<td>0.70 (0.59 to 0.81)</td>
<td>0.73 (0.61 to 0.84)</td>
<td>0.52 (0.26 to 0.82)</td>
</tr>
<tr>
<td>Management, κ (95% CI)</td>
<td>0.17 (-0.01 to 0.36)</td>
<td>0.48 (0.32 to 0.64)</td>
<td>0.54 (0.38 to 0.70)</td>
<td>0.32 (-0.07 to 0.75)</td>
</tr>
</tbody>
</table>

Table 3. Interrater Reliability of Main Outcomes

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Evaluator 2</th>
<th>Evaluator 3</th>
<th>Evaluator 4</th>
<th>Evaluator 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrater agreement by diagnosis for the primary differential diagnosis, % (n = 136)</td>
<td>46.6</td>
<td>NA</td>
<td>56.8</td>
<td>48.6</td>
</tr>
<tr>
<td>Interrater agreement by diagnostic category for the primary differential diagnosis, % (n = 136)</td>
<td>28.7</td>
<td>NA</td>
<td>43.2</td>
<td>50.0</td>
</tr>
<tr>
<td>Interrater agreement by management, % (n = 234)</td>
<td>31.6</td>
<td>NA</td>
<td>50.8</td>
<td>40.2</td>
</tr>
<tr>
<td>Diagnosis for the primary differential diagnosis, κ (95% CI)</td>
<td>0.41 (0.31 to 0.52)</td>
<td>NA</td>
<td>0.51 (0.41 to 0.61)</td>
<td>0.43 (0.34 to 0.53)</td>
</tr>
<tr>
<td>Diagnostic category for the primary differential diagnosis, κ (95% CI)</td>
<td>0.22 (0.14 to 0.31)</td>
<td>NA</td>
<td>0.37 (0.27 to 0.46)</td>
<td>043 (0.34 to 0.53)</td>
</tr>
<tr>
<td>Management, κ (95% CI)</td>
<td>0.08 (0.02 to 0.15)</td>
<td>NA</td>
<td>0.12 (0.01 to 0.23)</td>
<td>0.04 (-0.06 to 0.15)</td>
</tr>
<tr>
<td>Oral cases only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrater agreement by diagnosis for the primary differential diagnosis, % (n = 29)</td>
<td>66.7</td>
<td>68.0</td>
<td>66.7</td>
<td>61.5</td>
</tr>
<tr>
<td>Interrater agreement by diagnostic category for the primary differential diagnosis, % (n = 29)</td>
<td>34.6</td>
<td>64.0</td>
<td>66.7</td>
<td>64.0</td>
</tr>
<tr>
<td>Interrater agreement by management, % (n = 30)</td>
<td>32.4</td>
<td>44.4</td>
<td>46.2</td>
<td>34.6</td>
</tr>
<tr>
<td>Diagnosis for the primary differential diagnosis, κ (95% CI)</td>
<td>0.58 (0.39 to 0.79)</td>
<td>0.58 (0.36 to 0.80)</td>
<td>0.57 (0.36 to 0.78)</td>
<td>0.51 (0.33 to 0.73)</td>
</tr>
<tr>
<td>Diagnostic category for the primary differential diagnosis, κ (95% CI)</td>
<td>0.17 (0.04 to 0.36)</td>
<td>0.51 (0.32 to 0.74)</td>
<td>0.55 (0.37 to 0.76)</td>
<td>0.52 (0.32 to 0.73)</td>
</tr>
<tr>
<td>Management, κ (95% CI)</td>
<td>-0.01 (-0.21 to 0.13)</td>
<td>0.09 (-0.19 to 0.41)</td>
<td>-0.03 (-0.28 to 0.22)</td>
<td>-0.14 (-0.38 to 0.09)</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable.
determining the optimal amount of clinical experience that each mobile evaluator should have in working within the target patient population to achieve consistent remote diagnostic evaluations.

When looking at interrater reliability of our main outcomes, agreement was highest across all reviewers for diagnosis (47%-57%), with κ coefficients consistently in the moderate range (0.41-0.51) and even higher and more constant when the cases were limited to those with oral pathologic conditions (62%-68% agreement; κ range, 0.51-0.58). Thus, it seems that 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 evaluator 2, interrater reliability seemed fair to moderate for diagnostic category of the primary diagnosis overall and in the oral case subset. Diagnostic accuracy has been correlated with the degree of clinical experience.18 As noted earlier, evaluator 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 poorly correlated with the management choices of the face-to-face dermatologist. Mobile evaluators more frequently recommended management options that included further diagnostic testing (ie, test or test and treat), whereas the face-to-face dermatologist more frequently chose management options involving treatment alone due to technical limitations in the availability of diagnostic equipment (eg, biopsy kits) and the 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 interrater agreement among the mobile evaluators themselves. For the primary differential diagnosis, the mobile evaluators achieved moderate reliability for diagnosis with a κ coefficient of 0.44 (95% CI, 0.36 to 0.52), which increased to good reliability (κ, 0.61; 95% CI, 0.48 to 0.75) when limited to the cases with oral involvement. For diagnostic category of the primary differential diagnosis, the interrater reliability of the remote evaluators among themselves reached κ coefficients of 0.34 (95% CI, 0.27 to 0.42) overall and 0.37 (0.25-0.52) for the oral cases. Finally, agreement on management of the primary diagnosis was poor among the remote evaluators, with κ coefficients of 0.04 (95% CI, -0.04 to 0.12) for the dermatology-based evaluators and 0.01 (-0.13 to 0.18) among the oral evaluations. Broadly speaking, these levels of agreement are similar to the level of agreement each mobile evaluator achieved compared with the face-to-face analysis.

In terms of validity, with the exception of evaluator 3’s assessment of herpes simplex virus, specificity for the primary diagnosis outstripped sensitivity for the top 10 most common conditions identified by the face-to-face evaluator. These findings suggest that mobile teleconsultants may be better at ruling in the top 10 most common diagnoses in this population than they are at ruling them out (ie, 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 detecting malignant vs benign tumors by teledermatology was 0.99 (95% CI, 0.98-1.0), while specificity was 0.62 (95% CI, 0.56-0.69).19 Further research into which conditions are better suited for teledermatology consultation, particularly mobile teledermatology, is warranted.

In this pilot study, due to logistical limitations, the nurse interviewed and photographed patients after the face-to-face dermatologist more frequently chose management options involving treatment alone due to technical limitations in the availability of diagnostic equipment (eg, biopsy kits) and the 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 interrater agreement among the mobile evaluators themselves. For the primary differential diagnosis, the mobile evaluators achieved moderate reliability for diagnosis with a κ coefficient of 0.44 (95% CI, 0.36 to 0.52), which increased to good reliability (κ, 0.61; 95% CI, 0.48 to 0.75) when limited to the cases with oral involvement. For diagnostic category of the primary differential diagnosis, the interrater reliability of the remote evaluators among themselves reached κ coefficients of 0.34 (95% CI, 0.27 to 0.42) overall and 0.37 (0.25-0.52) for the oral cases. Finally, agreement on management of the primary diagnosis was poor among the remote evaluators, with κ coefficients of 0.04 (95% CI, -0.04 to 0.12) for the dermatology-based evaluators and 0.01 (-0.13 to 0.18) among the oral evaluations. Broadly speaking, these levels of agreement are similar to the level of agreement each mobile evaluator achieved compared with the face-to-face analysis.

In terms of validity, with the exception of evaluator 3’s assessment of herpes simplex virus, specificity for the primary diagnosis outstripped sensitivity for the top 10 most common conditions identified by the face-to-face evaluator. These findings suggest that mobile teleconsultants may be better at ruling in the top 10 most common diagnoses in this population than they are at ruling them out (ie, 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 detecting malignant vs benign tumors by teledermatology was 0.99 (95% CI, 0.98-1.0), while specificity was 0.62 (95% CI, 0.56-0.69).19 Further research into which conditions are better suited for teledermatology consultation, particularly mobile teledermatology, is warranted.
to-face evaluation was completed. Since 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 before the face-to-face evaluations.

Some of the differences we observed in interrater reliability of the main outcomes may be accounted for by the difficulty of defining certain conditions into 1 specific category. For example, while most of the mobile evaluators easily identified Kaposi sarcoma as the likely diagnosis in a given photo-case, they often differed when defining its diagnostic category, with some consistently choosing neoplasm and 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 a mobile teledermatology consultation, suggesting a diagnostic category might at least allow triage of the patient in terms of broad management steps (ie, biopsy, referral for a face-to-face dermatologic evaluation, or empirical treatment). Although the diagnostic categories that we listed have face validity (eg, were drawn from a gold-standard textbook of dermatology[20]) and had good intrarater 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 (ie, history and examination), the remote evaluators were restricted by at least 2 factors: a cellular phone interface that limited history intake to demographics, the use of yes/no or multiple-choice questions, and a data collector with limited knowledge of dermatology and dermatologic photography. For instance, when asked to rate the quality of the photographs they received (1, good; 2, satisfactory; or 3, poor), the mean photographic quality rating for most mobile evaluators fell between satisfactory and poor (mean test/retest values: evaluator 2, 2.1/2.3; evaluator 3, 2.5/2.6; evaluator 4, 1.4/1.4; and evaluator 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 cellular phone software will include the ability to incorporate free text, which will allow this hypothesis to be tested further. However, the data collection limitations imposed by a data collection intermediary with limited dermatologic knowledge will likely remain; most health care workers seeking mobile teledermatology consultations in the real-world setting are not specifically trained or equipped to take photographs in a way that is ideal for mucocutaneous evaluation. For example, even with cellular phones equipped with high-quality cameras, photographs 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 toward 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 photographs in the real-world setting may alleviate some of these limitations by improving the quality of the data that are captured.

Conclusions

As evidenced by our pilot study, although the introduction of mobile teledermatology into a resource-limited population of HIV-positive patients in sub-Saharan Africa has significant theoretical potential for improving access to care, much work is 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.

ARTICLE INFORMATION

Accepted for Publication: August 2, 2013.

Author Contributions: Drs Azfar and Kovarik had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Azfar, Bilker, Gelfand, Kovarik.
Acquisition of data: Azfar, Lee, Castelo-Soccio, Greenberg, Kovarik.
Analysis and interpretation of data: Azfar, Bilker, Gelfand, Kovarik.

Obtained funding: Azfar, Kovarik.
Administrative, technical, or material support: Azfar, Kovarik.

Conflict of Interest Disclosures: Dr Gelfand reported receiving grants from Amgen, Pfizer, Novartis, and Abbott and is a consultant for Amgen, Abbott, Pfizer, Novartis, Celgene, and Centocor. No other disclosures were reported.

Affiliations: Department of Medicine, University of California, San Francisco, CA (Azfar, Kovarik); Research and Development, Grammologics, Inc, Mountain View, CA (Azfar); Department of Dermatology, University of California, San Francisco, CA (Kovarik); Department of Dermatology, University of California, San Francisco, CA (Lee, Castelo-Soccio); Department of Dermatology, University of California, San Francisco, CA (Greenberg).

Copyright 2014 American Medical Association. All rights reserved.
Mobile Teledermatology in HIV-Positive Patients

Funding/Support: This work was supported by grants F32-AR056799 (Dr Azfar) and K24-AR064310 (Dr Gelfand) from the National Institute for Arthritis and Musculoskeletal and Skin Diseases (NIAMS), a grant from the Center for Public Health Initiatives (Dr Azfar and Kovarik), funding from the Center for AIDS Research through the University of Pennsylvania (Dr Kovarik and Azfar), and dermatology departmental training grant T32 from the University of Pennsylvania (Dr Castelo-Soccio).

Role of the Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Worship Muzangwa, RN, Gordana Cavric, MD, PhD, and Zola Musimar, MB, BCh, at the Princess Marina Hospital; Diana Dickinson, MD, at the Independence Surgery Centre; and colleagues at the Princess Marina Hospital and the Athlone Hospital, the Infectious Disease Care Centre, the Botswana-UPenn Partnership, the Ministry of Health of the Government of Botswana, ClickDiagnostics, and the Medical University of Graz provided logistical assistance in implementing this study.

REFERENCES


