Inter-observer variability of teledermoscopy.

Marius Rademaker, Eugene Tan, Amanda Oakley
Dermatology Department, Waikato Hospital, Hamilton, New Zealand

Abstract. To assess the inter-observer variability between teledermoscopists, images from 979 lesions from 206 New Zealand patients were distributed to 5 independent experienced teledermoscopists in New Zealand, Australia and America.

There was excellent agreement between 4 out of 5 teledermoscopists for lesions that were agreed upon as melanoma. The fifth dermatologist made a more frequent diagnosis of melanoma than the others. The agreement for melanocytic lesions was better than that for non-melanocytic lesions: mean kappa values of agreement for benign naevus was K=0.80 and atypical naevus K=0.75 whereas seborrhoeic keratosis K=0.72, BCC K=0.61, solar keratosis K=0.42, SCC-in-situ K=0.24, and invasive SCC K=0.10. There was good ability to distinguish malignant from benign lesions particularly for basal cell carcinoma.

The Australian/New Zealand dermatologist had good agreement in the diagnosis of melanoma and atypical naevi, which reflects their degree of comfort with their own patient populations. The Northern hemisphere dermatologist made a diagnosis of melanoma and atypical naevi more frequently, which maybe explained by lack of familiarity with the specific study population. This difference also reflects the lack of consensus guidelines in definition of atypical naevus and possibly diagnostic drift.

Introduction

Teledermatology has evolved through the decades and has been used for various purposes ranging from triage to diagnostics. Teledermoscopy adds epiluminescence microscopy or dermatoscopy to further increase diagnostic accuracy. The purpose of this study was to assess the inter-observer variability in teledermoscopy for the diagnosis of lesions from patients referred by their general practitioner to a lesion diagnosis clinic. External validity was increased by recruiting experienced teledermoscopists from Australia, New Zealand and the United States of America.

Methods

Patients referred to a hospital specialist skin lesion clinic for diagnosis and management of one or more skin lesion(s) were recruited. Panoramic views of the body were first taken to map the location of the lesion(s); followed by macroscopic views (30mm field of view, 'macro') and then dermoscopic views (15mm field of view, 'micro') of the lesion(s). A proprietary software (MoleMap Point of Diagnosis software developed by MoleMap, Unit L/383 Khyber Pass Road, Newmarket, Auckland, New Zealand), was used to manage these files via a remote server, allowing the dermatologists to analyse the data in any location using an authorised computer.

A total of 979 lesions were obtained and the database was distributed to five experienced teledermoscopists, two in New Zealand (A and B), two in Australia (C and D) and one in United States of America (E).

Lesions that were diagnostically agreed upon by all 5 teledermoscopists were considered the standard of reference to which inter-observer variability was measured. Dermatologist B used the term 'atypical naevi requiring excision' for lesions suspect of being melanoma: for the purposes of analysis, these patients were considered in the melanoma group. Pathological data was not used as a comparison as the aim of this study was to assess inter-observer variability, and not diagnostic accuracy, which has been assessed in many earlier trials.

Kappa values were calculated using Stata software (Stata Press, College Station, TX version) to assess the proportion of inter-observer agreement beyond that expected by chance between pairs of observers. Perfect agreement is indicated by a kappa value of 1.0, whereas a kappa value of 0 indicates no agreement at all. We distinguish between the following levels of agreement between observers for the indicated kappa values: Poor agreement, <0.20; slight agreement, 0.21–0.40; moderate agreement, 0.41–0.60; substantial agreement, 0.61–0.80; and almost perfect agreement, > 0.81 between observers.

Results

Table 1 shows the distribution of the diagnoses made by the five teledermoscopist. The most common malignant lesion was basal cell carcinoma where up to 80 lesions were diagnosed. The frequency of melanoma diagnosis ranged from 10 to 61. The most common benign diagnoses were benign naevus and seborrhoeic keratosis. The category of "other" included a miscellaneous group of skin conditions including psoriasis, chondrodermatitis nodularis helicis, insect bites, cysts, granulomas, viral warts and sebaceous hyperplasia. Teledermoscopist A-D were uncertain about 2-6 lesions whilst E was uncertain about 42 lesions.

	A	B	C	D	E
Malignant lesions					
Melanoma	16	10*	22	17	61
BCC	70	61	64	80	73
SCC/KA	24	17	39	12	22
Benign lesions					
SCC-in-situ	24	35	6	11	58
Solar Keratosis	79	97	80	82	39
Seborrhoeic Keratosis	109	89	121	110	92
Benign Naevus	588	592	579	604	92
Atypical Naevus	26	39	28	44	440
Dermatofibroma	5	1	3	3	6
Haemangioma	11	12	13	10	7
Other	21	23	24	4	47
Uncertain	6	3	0	2	42

Table 1. Diagnosis made by 5 dermoscopist (A-E).

Teledermoscopist A-D showed substantial degrees of diagnostic agreement amongst one another for the diagnosis of melanoma, with K value > 0.80. Teledermoscopist E varied from the group with the

diagnosis of 46 additional melanomas and had a kappa value to 0.38.

	A	B	C	D	E
Malignant lesions	0.84	0.93	0.74	0.85	0.57
Benign lesions	0.46	0.09	0.14	0.07	0.7
Melanoma	0.97	0.85	0.81	0.94	0.38
Benign Naevus	0.10	0.10	0.11	0.10	0.89
Atypical Naevus	0.14	0.09	0.13	0.08	0.
Seborrhoeic Keratosis	0.69	0.80	0.64	0.69	0.78
SCC-in-situ	0.32	0.22	0.3	0.2	0.15
Solar Keratosis	0.38	0.32	0.38	0.37	0.67
BCC	0.61	0.67	0.65	0.55	0.59
Invasive-SCC	0.08	0.11	0.05	0.15	0.09

Table 2. Kappa values

Teledermoscopist E diagnosed a greater number of additional malignant lesions (109 malignant lesions) above that which was agreed upon by the group. This translates to the lowest kappa value for agreement of 0.57 (Table 2). Accordingly, when kappa statistics were calculated for benign lesions, teledermoscopist E appeared to have the highest kappa score due to the lowest number of additional lesions or lowest deviation from the group. However, when teledermoscopist E was excluded from the analysis, kappa values for the other 4 teledermoscopists rose significantly. Teledermoscopist E also varied significantly from the others in the benign category by diagnosing a greater number of atypical naevi (440 lesions) and fewer benign naevi (92 lesions).

There was more variability in the diagnosis of non-melanocytic lesions, particularly for benign and pre-cancerous lesions such as SCC-in-situ and solar keratosis where the agreement was only slight to moderate (Table 2). There was poor agreement for lesions that were considered invasive-SCC.

Discussion

This study has shown that for melanoma, the diagnostic agreement was generally excellent. It complements the findings from earlier studies on the diagnostic accuracy of teledermatology and teledermoscopy.

The increased diagnosis of melanoma and atypical naevi by one of the teledermoscopist raises important issues for discussion. Several explanations could account for this variability. Firstly, it underscores the difficulty and vagueness in the definition of atypical naevus. It has been defined in the literature as a "naevus with architectural disorder". Despite a consensus conference in 1992, wide variations in the terminology still persist. This variability could be explained by differences in clinical judgment, which is a complex and subjective process.

Familiarity with the study population may play a significant role. The population studied were individuals who were all from 1 center in New Zealand. Different populations display consistent differences in colour, shape and dimension between their naevi, adding complexities to the definition of atypical. These variations could be accounted for by differences in genetic make-up and environmental factors such as ultraviolet exposure. The Australasian dermatologists (A-D) showed greater consistency in this respect, reflecting their experience of this particular population of naevi.

In addition, clinical judgement and diagnostic threshold is dependent on feedback, both positive and negative, from pathological reports. Pathologists in different parts of the world may over- or under-diagnose melanoma and atypical naevi, which further blurs the distinction between lesions that are classified as atypical/dysplastic versus melanoma. Although pathological diagnosis is often cited as the "gold standard", several reports have documented significant variability amongst pathologists in the diagnosis of melanoma and even more so for SCC. This tendency to either over- or under-diagnose melanoma could lead to a pathological induced 'drift' in clinical diagnosis by the dermatologist. The litigious environment of the USA (or lack of it in New Zealand) may also play a role in diagnostic drift.

Summary

In summary, teledermoscopy shows good agreement for melanoma amongst 4 out of the 5 dermatologists but variability exists between Northern and Southern hemisphere dermatologists. One of the five teledermoscopists made a diagnosis of melanoma and atypical naevus more frequently, a difference that is likely explained by the lack of familiarity with the specific patient population, lack of consensus guidelines in definition of an atypical naevus and/or pathologic induced clinical diagnostic drift. There was also more variability in the diagnosis of non-melanocytic lesions.

Inter-observer variability amongst dermatologists is an important aspect for determining clinical accuracy and should be the focus of further research.