

# Reliability of Computer Image Analysis of Pigmented Skin Lesions of Australian Adolescents

Joanne F. Aitken, Ph.D.<sup>1</sup>  
 John Pfitzner, M.App.Sc.<sup>1</sup>  
 Diana Battistutta, B.Sc.<sup>1</sup>  
 Peter K. O'Rourke, Ph.D.<sup>2</sup>  
 Adèle C. Green, M.B.B.S.<sup>1</sup>  
 Nicholas G. Martin, Ph.D.<sup>1</sup>

<sup>1</sup> Queensland Institute of Medical Research, Brisbane, Australia.

<sup>2</sup> Department of Social and Preventive Medicine, University of Queensland, Brisbane, Australia.

**BACKGROUND.** The diagnosis of melanomas at an early stage is associated with improved survival, so the recognition of changes in pigmented skin lesions over time is important. We have developed a computer imaging system with the aim of assisting clinicians in differentiating early melanomas from benign pigmented skin lesions. The objective of this study was to investigate the system's reliability over time in measuring diagnostic characteristics of pigmented skin lesions, including their color, size, shape, and distinctness of boundary.

**METHODS.** We captured video images of 5 lesions, all larger than 2 mm in greatest dimension, on each of 66 Australian adolescents on 2 occasions approximately 1 month apart. Features extracted by computer image analysis included area, perimeter, and regularity of outline of the lesions, the mean and standard deviation of reflectance at red, green, and blue wavelengths, and the mean and standard deviation of the gradients of red, green, and blue reflectance at the lesion boundary.

**RESULTS.** All measurements showed moderate to high reliability (intraclass correlation coefficients 0.66–0.94), except for the standard deviations of the color gradients, whose reliability improved to moderate levels (0.68–0.71) when the mean of 5 lesions was considered. For most outcomes, reasonable within subject reliability was achieved when five lesions per subject were measured.

**CONCLUSIONS.** These results, in combination with previous work demonstrating the reasonable ability of this computer imaging system to discriminate between malignant melanomas and other pigmented lesions, indicates that the system has the potential to become a useful tool for clinicians in following people with pigmented lesions over time to detect early malignant changes. *Cancer* 1996; 78:252–7. © 1996 American Cancer Society.

**KEYWORDS:** melanoma, image analysis, early diagnosis, skin cancer, nevi.

The recognition and excision of melanomas at an early stage is associated with prolonged disease free survival,<sup>1</sup> and consequently the assessment of the malignant potential of pigmented skin lesions has become part of routine medical care.<sup>2</sup> This is particularly so in Queensland, where malignant melanoma is one of the leading causes of cancer death. Further, as the presence of melanocytic nevi is the strongest known predictor of melanoma risk, recognizing and counting pigmented skin lesions of various types is an integral part of epidemiologic investigations of melanoma etiology. Unfortunately, differentiating early melanomas, melanocytic nevi, freckles, keratoses, and other pigmented lesions accurately and consistently is a difficult task even for trained observers. To address this problem we are developing a computer imaging system for characterizing and counting pigmented skin lesions, with eventual application in the clinical setting.

In a digital image of the skin surface, lesions appear as isolated clusters of picture elements (pixels) which differ in color and bright-

Supported by the Queensland Cancer Fund, and Grant No. 950998 of the National Health and Medical Research Council of Australia.

The authors thank Bruce McGregor for critical comments on an earlier draft of this manuscript.

Address for reprints: Joanne F. Aitken, Ph.D., Queensland Institute of Medical Research, P.O. Royal Brisbane Hospital, Brisbane, Australia 4029.

Received April 1, 1996; accepted April 1, 1996.

ness from those of neighboring regions of skin. With the appropriate computer program, it is thus possible to identify and label the position of the lesions within the image, and make quantitative measures of their distribution, color, and geometry.<sup>3-5</sup> At present, our system is in two parts: a standard commercial video camera records macroscopic images of the whole back, and a separate minivideo camera records images of individual lesions. Recently, we have shown that counts of total pigmented lesions on the back by macroscopic image analysis are reproducible, with an intraclass correlation coefficient of 0.92 for repeated counts approximately 1 month apart in a sample of 66 adolescents.<sup>6</sup> We have also demonstrated that the microimaging system can discriminate to some extent between melanomas and benign pigmented lesions on the basis of their size, shape, and color, with an overall accuracy of 89%.<sup>7,8</sup>

The broad aim of this study was to evaluate further the performance of computer image analysis by assessing the reliability (repeatability) of measurements of area, perimeter, regularity of outline, color variation, and distinctness of boundary with surrounding skin obtained from images of pigmented skin lesions. Our objectives were two-fold: first, to estimate the reliability over time of measurements of individual lesions (within lesion reliability); and second, to estimate the reliability over time of the mean of five lesions per subject (within subject reliability). The first is of most interest in clinical and diagnostic applications, in particular, in following pigmented lesions over time to detect early malignant changes. The second is of interest in epidemiologic studies, where the focus is on differences between individuals in the average size and other characteristics of their pigmented lesions, and the genetic and environmental causes of such differences. In particular, in twin studies where the goal is to estimate the proportion of variance due to genetic factors (heritability), the within subject reliability must represent an upper limit to the heritability.<sup>9</sup>

## METHODS

### Subjects and Clinical Examinations

The study was conducted as part of an ongoing longitudinal investigation of nevus development in adolescent twins in Queensland. The sample and clinical protocol have been described elsewhere.<sup>6</sup> Briefly, twins 10 to 12 years were ascertained from primary schools in Brisbane and surrounding areas and invited to take part in the study. So far, 306 pairs have taken part with parental consent. Among these, 33 pairs of twins age 12 years (all white, 50% male) were examined twice early in the study and are the subjects of this

analysis. As the focus is on reliability of measurement, twin pairing was ignored for this analysis.

Each subject was examined at the Queensland Institute of Medical Research on 2 separate occasions at least 2 weeks apart (range: 2–11 weeks; median: 4 weeks). On each occasion, a video camera was used to record a large-scale image of the back, while a second minivideo camera was used to obtain closeup images of 5 lesions over 2 mm in greatest dimension, randomly selected from all lesions on the body which were greater than 2 mm. The total number of lesions greater than 2 mm per subject ranged from 4 to 249, with a median of 33 such lesions per subject.<sup>6</sup> Each of the five lesions was filmed twice, removing the camera between each sequence. Each selected lesion was numbered with an adhesive label and captured by macrovideo image for accurate identification at the second visit, when the same five lesions were again filmed twice.

### Computer Imaging System

The computer imaging protocol has been described in detail elsewhere.<sup>8</sup> The image recording equipment consists of a miniature color video camera (GEC Panasonic, Matsushita Electrical Industrial Co., Ltd., Central P.O. Box 288, Osaka 530-91, Japan) with built-in color standards and an annular fibre-optic light source (SCHOTT, Glaswerker, works Wiesbaden Product Group, Fiber Optics, P.O. Box 130367 D-6200 Weisbaden 13) mounted in a hand-held device. The device has a 2 cm square aperture which is placed over the lesion, thus ensuring constant focus and illumination. An 8-second segment of videotape was recorded for each lesion on an SVHS video cassette recorder (GEC Panasonic, Matsushita Electrical Industrial Co., Ltd., Central P.O. Box 288, Osaka 530-91, Japan), and subsequently red, green, and blue images were captured from the replayed videotape by a color frame grabber (Vision Plus-AT Imaging Technology Inc., 600 West Cummings Park, Woburn, Massachusetts 01801) mounted in an IBM-compatible 386 Microcomputer (NEC Corporation, Tokyo, Japan); 4 frames of each lesion were captured. This protocol had been configured and tested previously during several related projects in our laboratory. This entire process was performed twice for each lesion at each of the two visits. Each 24-bit color image (8 bits each for red, green, and blue reflectance) was at a spatial resolution of  $512 \times 512$  pixels. Digitized images were archived to tape, and analysis software was used to extract the required features from the images on a HP 9000 Workstation (Hewlett Packard, 3000 Hanover Street, Palo Alto, California 94304). The lesion boundary was determined using a color space thresholding algorithm and an

**TABLE 1**  
**Estimates of Components of Variance for Measurements of the Size, Shape, and Color of Pigmented Skin Lesions on 66 Australian Adolescents Using Computer Image Analysis<sup>a</sup>**

Measurements by computer image analysis	Components of variance			
	Subject	Lesions within subject	Visit $\times$ lesions within subject	Error
Area	38.7	57.7	0.5	3.1
Area change	23.5	61.2	3	12.2
Perimeter	34.6	59.8	0.5	5.1
Fragmentation index	11.7	63.4	3.8	21.1
Mean red intensity	41.9	51.4	3.4	3.3
Mean green intensity	39.7	52.6	4	3.7
Mean blue intensity	41	50.5	4.2	4.3
Mean SD of red intensity	25.5	63.1	2.2	9.2
Mean SD of green intensity	21.5	59	3	16.5
Mean SD of blue intensity	24.9	49.5	3.2	22.4
Mean gradient of red intensity	26.8	60.4	0.9	11.9
Mean gradient of green intensity	24.1	57.3	0.7	17.9
Mean gradient of blue intensity	21.4	58.4	0.9	19.4
Mean SD of gradient of red intensity	22	24.7	2.9	50.4
Mean SD of gradient of green intensity	19.2	20.3	4.4	56.1
Mean SD of gradient of blue intensity	20.4	20.1	4.7	54.8

<sup>a</sup> Components of variance are standardized to 100% for ease of comparison.

SD: standard deviation.

eight neighborhood contour follower (described in detail elsewhere).<sup>8</sup>

Measurements were extracted from each frame and averaged across the four frames of each lesion. The measurements were the area of the lesion and its perimeter; mean reflectance at the red, green, and blue wavelengths within the lesion boundary; the standard deviations of red, green, and blue reflectance (giving an indication of color variation within the lesion); and the means and standard deviations of the gradient in red, green, and blue reflectance at each boundary pixel (the gradient at each boundary pixel was determined from the averages of skin and lesion pixels in a 25 neighborhood mask).<sup>8</sup> We also calculated the change in area when the color thresholds were reduced by 10% as an indication of the gradient of color change around the lesion boundary; and fragmentation index ( $4\pi$  area/perimeter<sup>2</sup>) as a measure of the regularity of the border. Fragmentation index relates the perimeter of the lesion to a circle as a standard, taking a value of 1 if the lesion is a perfect circle, with smaller values for more fragmented lesions.

### Statistical Analysis

Five lesions on each of 66 subjects imaged twice during each of 2 visits gave a possible total of 1,320 mea-

surements for each variable (66 subjects  $\times$  5 lesions  $\times$  2 visits  $\times$  2 images). Due to contrast or logistic problems, 9 subjects had a total of 29 missing measurements, leaving 1,291 measurements for analysis. For most of the variables, measurements were transformed using the square root transformation to stabilize variances and improve normality. Exceptions were the fragmentation index for which an arctan transformation was used, and the means of red, blue, and green reflectance, which were not transformed.

For all variables, the intraclass correlation coefficient was used as the measure of agreement (reliability) between readings at the first and second visits, with 0 representing no agreement and 1 perfect agreement between readings on the two occasions.<sup>10</sup> Estimation of these coefficients was based on analysis of variance components.<sup>11</sup> Variance components were estimated using a restricted maximum likelihood approach using the VARCOMP procedure in SAS (UNIX),<sup>12</sup> with subjects and lesions within subjects considered to be random effects and visits a fixed effect. Repeated images of each lesion at each time were treated as replicates in the design; the resulting model allowing derivation of variance components for subjects, lesions within subjects, visits by lesions within subjects, and error.

**TABLE 2**  
Intraclass Correlation Coefficients for Reliability over Two Occasions of Measurements of Pigmented Skin Lesions on 66 Australian Adolescents Using Computer Image Analysis

Measurements by computer image analysis	Intraclass correlation coefficients for reliability of measurements of	
	Individual lesions	Mean of five lesions per subject
Area	0.94	0.76
Area change	0.80	0.62
Perimeter	0.92	0.74
Fragmentation index	0.72	0.43
Mean red intensity	0.88	0.78
Mean green intensity	0.87	0.76
Mean blue intensity	0.86	0.77
Mean SD of red intensity	0.85	0.64
Mean SD of green intensity	0.75	0.60
Mean SD of blue intensity	0.66	0.66
Mean gradient of red intensity	0.83	0.67
Mean gradient of green intensity	0.76	0.65
Mean gradient of blue intensity	0.74	0.62
Mean SD of gradient of red intensity	0.32	0.71
Mean SD of gradient of green intensity	0.25	0.68
Mean SD of gradient of blue intensity	0.25	0.69

SD: standard deviation.

Two intraclass correlation coefficients were computed for each variable. First, the reliability of measurements of individual lesions was computed as

$$\sigma_{l:s}^2 / (\sigma_{l:s}^2 + \sigma_{v^*l:s}^2 + \sigma_e^2),$$

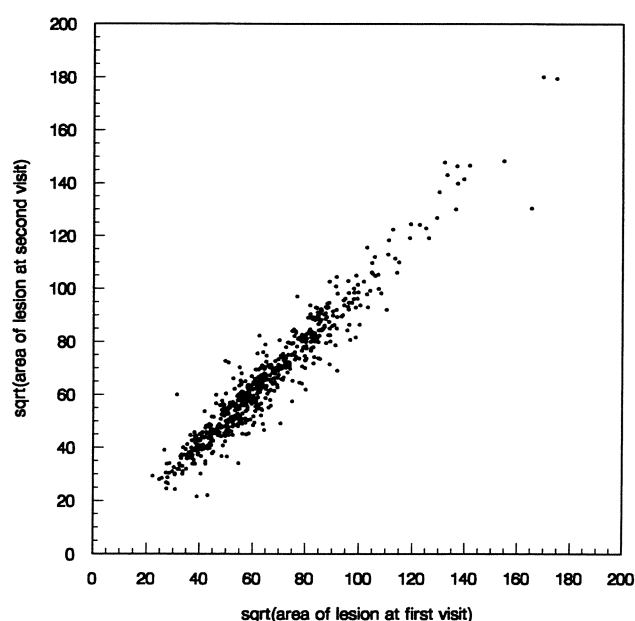
where the subscript  $l:s$  refers to lesions within subjects,  $v^*l:s$  to visits by lesions within subjects, and  $e$  to residual measurement error. Second, the reliability of the mean of five lesions per subject was computed as

$$\sigma_s^2 / (\sigma_s^2 + \sigma_{l:s}^2/5 + \sigma_{v^*l:s}^2/2 + \sigma_e^2/20),$$

where  $\sigma_s^2$  refers to the component of total measurement variation attributable to study subjects. We also considered how the reliability of the mean of multiple lesions per subject increased with increasing numbers of lesions per subject.<sup>9</sup>

## RESULTS

For area, perimeter, and the mean color intensities most of the measurement variation was due to differences between subjects and between lesions within subjects, with random measurement error contribut-

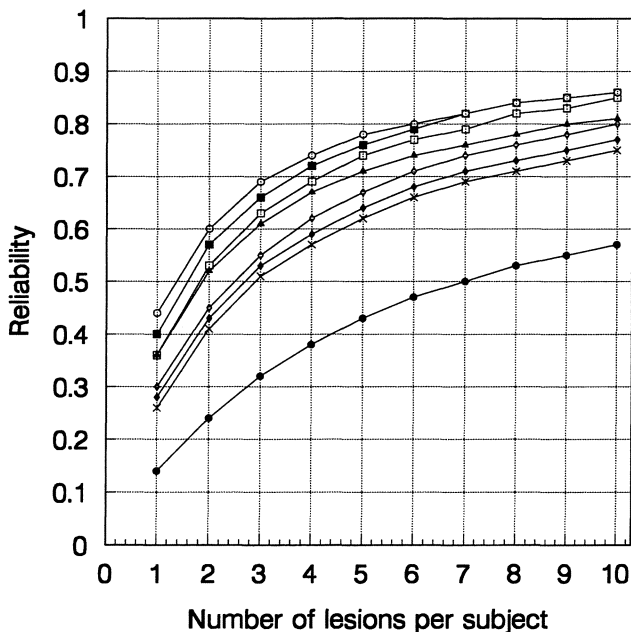


**FIGURE 1.** Area of pigmented skin lesions on 66 Australian adolescents (5 lesions per subject) measured using computer image analysis on 2 occasions approximately 1 month apart (intraclass correlation coefficient between measurements on the two occasions = 0.94) are shown. Area measurements were transformed by the square root transformation to improve normality.

ing  $\leq 5.1\%$  of the total variation (Table 1). Random measurement error was higher for the remaining variables, although subjects and lesions remained the dominant sources of variation for all but the standard deviations of the color gradients. For each variable, the interaction of visits with lesions made only a small contribution to total variance, indicating that effects were not specific to visits.

The reliability of measurements of individual lesions between the 2 visits was greatest ( $\geq 0.80$ ) for area, area change, perimeter, the color intensity means, the red standard deviation, and the red gradient (Table 2, Fig. 1). All other outcomes showed moderate reliability (0.66–0.79), except for the color gradient standard deviations.

The reliabilities of the mean of five lesions per subject (within subject reliability) were all at moderate levels (0.60–0.78), with the exception of fragmentation index (Table 2). Figure 2 compares within subject reliability for several of the computer image outcomes, as a function of the number of lesions sampled per subject. For almost all outcomes, most of the increase in reliability relative to one lesion was achieved when five lesions per subject were measured. The exception was fragmentation index, where within subject rela-



**FIGURE 2.** Reliability of the mean of measurements by computer image analysis of pigmented skin lesions, according to the number of lesions sampled per subject is shown (■ = area, X = area change, □ = perimeter, ● = fragmentation index, ○ = mean red intensity, ◆ = mean standard deviation of red intensity, ◇ = mean gradient of red intensity, ▲ = mean standard deviation of gradient of red intensity).

bility of 0.43 with 5 lesions increased to 0.57 if 10 lesions were measured.

## DISCUSSION

We have described a computer imaging system currently under development for eventual application as an aid to assisting clinicians in diagnosing early melanoma, and in counting pigmented lesions including melanocytic nevi in high risk patients. Here we have examined the reproducibility of measurements by computer image analysis of the size, shape, and color of pigmented lesions, parameters used in clinical practice to discriminate melanoma from other pigmented lesions.<sup>13</sup> Our results indicate that most of the variation in repeated measurements by computer image analysis of the area, perimeter, shape, and color of individual pigmented skin lesions is explained by differences between subjects and between lesions, and that measurements are highly reproducible between separate visits. Further, these results show that for epidemiologic applications in which the focus is on differences between individuals satisfactory within subject reliability for most variables is obtained with five lesions per subject, and little is achieved by measuring more lesions.

Computer imaging, while not yet in routine use in clinical dermatology, is a rapidly expanding field with potential applications in research and clinical practice, particularly in screening for and follow-up of early malignant melanoma. Recent reviews have discussed current research activities in computer imaging for dermatology,<sup>14</sup> and the possible future of this technology as an aid for clinicians in early melanoma detection.<sup>15,16</sup> One of the potentially most useful applications of computer imaging is in following changes in pigmented lesions over time to observe their natural history and to signal early malignant changes.<sup>17</sup> In both of these applications, the differences between sequential measurements are of interest, thus the primary requirement is that the system demonstrate a high level of reproducibility of measurements between visits. The present results show that the imaging system we have developed meets this criterion.

In previous work, 89% of 164 unselected pigmented lesions, including 16 of 18 histologically confirmed melanomas, were correctly classified as melanoma or "other pigmented lesions" using discriminant analysis based on image analysis of measurements collected with the microimaging system described above.<sup>8</sup> The features which allowed the best discrimination between the two classifications were the size of a lesion, measured by area and perimeter, and its color variation, measured by the color standard deviations (indicating color variation within the lesion) and the color gradient standard deviations (indicating color variation at the lesion boundary). In the present study, all of these measurements except the last demonstrated moderate to high reliability. The color gradient standard deviations were less reliable than the other measurements, possibly because these were highly dependent on the precise lesion boundary defined by the thresholding algorithm at each visit, which itself was based on color change.

We have also shown previously that counts of pigmented lesions on the backs of Australian adolescents are highly reliable using the macroimaging system, with an intraclass correlation coefficient of 0.92 for independent counts on two occasions approximately 4 weeks apart,<sup>6</sup> suggesting that the system has a potential epidemiologic research application. In combination with our demonstration of the reasonable ability of the system to discriminate between malignant melanomas and other pigmented lesions,<sup>7,8</sup> the present results indicate that the microimaging system also has the potential to become a useful tool for both clinicians and epidemiologists in recognizing and documenting individual pigmented skin lesions over time.

## REFERENCES

1. Koh HK. Cutaneous melanoma. *New Engl J Med* 1991; 325:171-82.
2. Fleming WB. The cancer-related health check-up. A guide for medical practitioners. *Med J Aust* 1985;143 (Suppl):34S-40S.
3. Castleman KR. Digital image processing. Englewood Cliffs (NJ) Prentice-Hall, 1979.
4. Gonzalez RC, Woods RE. Digital image processing. New York: Addison-Wesley, 1993.
5. Russ JC. The image processing handbook. 2nd ed. Boca Raton: CRC, 1995.
6. Aitken JF, Green A, Eldridge A, Green L, Pfitzner J, Battistutta D, et al. Comparability of naevus counts between and within examiners, and comparison with computer image analysis. *Br J Cancer* 1994;69:487-91.
7. Green A, Martin N, McKenzie G, Pfitzner J, Quintarelli F, Thomas BW, et al. Computer image analysis of pigmented skin lesions. *Melanoma Res* 1991;1:231-6.
8. Green A, Martin N, Pfitzner J, O'Rourke M, Knight N. Computer image analysis in the diagnosis of melanoma. *J Am Acad Dermatol* 1994;31:958-64.
9. Falconer DS. Introduction to quantitative genetics. London: Longman, 1981:128-30.
10. Snedecor GW, Cochran WG. Statistical methods. 7th ed. Iowa: Iowa State University Press, 1980:234.
11. Dunn G. Design and analysis of reliability studies: the statistical evaluation of measurement errors. London: Oxford University Press, 1989:129-33.
12. SAS/STAT User's Guide, Version 6. 3rd ed. Cary (NC): SAS Institute Inc., 1990.
13. NIH Consensus Conference. Diagnosis and treatment of early melanoma. *JAMA* 1992;268:1314-8.
14. Stoecker WV, Moss RH. Digital imaging in dermatology [Editorial]. *Comput Med Imaging Graph* 1992;16:145-50.
15. Hall PN, Claridge E, Morris Smith JD. Computer screening for early detection of melanoma—is there a future? *Br J Dermatol* 1995;132:325-38.
16. Perednia DA. What dermatologists should know about digital imaging. *J Am Acad Dermatol* 1991;25:89-108.
17. Voight H, Claßen R. Topodermatographic image analysis for melanoma screening and the quantitative assessment of tumor dimension parameters of the skin. *Cancer* 1995;75:981-8.