Chapter 4: Validation experiments

4.1 Introduction

This study represents the first clinical trial with the SIAscope, a system that produces information about the haemoglobin, total melanin, dermal melanin and collagen content of the epidermis and papillary dermis within the region of interest scanned. Studies have been performed that measured the theoretical accuracy of the system in determining these parameters [Cotton, 1998; Hojjatoleslami *et al.*, 2000]. It was decided that experiments should be undertaken that could determine whether the SIAscope was indeed measuring these parameters. The four sets of experiments determining each of the SIAscope parameters are described below in the style of a short paper.

4.2 Total Melanin & Haemoglobin

4.2.1 Methods

In order to test these SIAscopy parameters, experiments were devised that were based on measurements taken from the dorsal and volar skin of the forearms of normal healthy volunteers. Normal skin contains only epidermal melanin and areas of uniform pigmentation without naevi or other lesions were sampled. The experiment attempted to correlate the SIAscope measurements with those of the Mexameter (CK Electronic, Cologne, Germany). The Mexameter is an industry standard spectrophotometer designed to measure melanin and haemoglobin content in the skin. It is mainly used by the cosmetic industry to measure erythema as an index of skin irritation and allergic reaction. In addition, melanin content is used as a measure of tanning response and to determine skin phototype. The device emits light over 3 wavelengths, namely 568, 660 & 880nm, and measures remitted light over a 5mm diameter. The erythema and melanin indexes are determined as follows:

$$Mx = \frac{500}{\log 5}.\log\frac{Infrared}{\mathrm{Re}\,d} + \log 5 \qquad (4.1)$$

$$Ex = \frac{500}{\log 5}.\log\frac{\mathrm{Re}\,d}{Green} + \log 5 \qquad (4.2)$$

(where Mx = melanin index, Ex = erythema index & infrared/red/green = infrared/red/green remittance)

These indices are relative values and the maximum ratio between each colour is 1:5. The range of values is 0-1000, with a higher value representing more melanin or erythema, and a value of 500 represents a remittance ratio of 1:1 [Mexameter technical manual].

There is an obvious size discrepancy between the imaging handsets so it was decided that taking the average of 3 readings using the Mexameter and an average of the melanin and haemoglobin values in the central 10mm square of the total melanin SIAgraph should assess the melanin content.

The hypothesis was that an increase in melanin or haemoglobin concentration as measured by the Mexameter (the standard for this experiment) would be matched in melanin or haemoglobin concentration as determined by the SIAscope. This is a clear case for linear regression analysis [Campbell & Machin, 1999]. In cases where two devices are being compared that measure the same parameter, such as blood glucose concentration, then the statistic described by Bland & Altman (1986) is appropriate. In short, this statistic assumes that if the measurements being produced by the two systems are the same then the differences in the values between the systems will remain constant throughout a range of values of the parameter being measured. However, in this experiment this statistic is not appropriate as the SIAscope and Mexameter use different units of measurement. Thus, as the melanin concentration increases, so does the difference in the measurements between the two systems and the assumptions of the Bland & Altman statistic are violated.

In linear regression problems we are interested in how well one variable can be used to predict another. Thus with this experiment we are concerned with predicting the melanin concentration, indirectly determined using the Mexameter, from the SIAscope measurement. The Mexameter reading is known as the outcome variable and the SIAscope measurement is known as the predictor or explanatory variable. Alternatively the Mexameter measurement can be described as the **dependent** variable and the SIAscope measurement can be described as the **independent** variable. It is the latter terminology that will be used in this section. If the independent-dependent pairs of values were plotted on a graph, with the independent values classically on the X-axis, then it would be hoped that the scatter-plot would reveal a linear trend. In a simple linear relationship we can describe this relationship by the equation:

$$Y = a + bX + E \tag{4.3}$$

where Y is the dependent variable, X is the independent variable, a and b are constants and E is a random variable, with a mean of zero, that represents the part of the variability of Y that is not explained by the relationship with X. This variable is called the error. The constant, *a*, is the intercept, the value of Y where X is zero, and the constant b represents the change in Y per unit value change in X. A scatter-plot of all X-Y pairs could have many lines drawn through it to

describe the data, but the line that best describes the data will be the one that has the minimum amount of variability of Y that is unexplained. This means that the line of best fit is one that has the minimum variance of E and can be calculated by the method of least squares.

Once the line of best fit has been determined it is necessary to determine the *goodness-of-fit* of this line to the data. SPSS for Windows (SPSS Inc) produces 3 basic measurements based on the sum of squares that gives the experimenter an idea of the contribution of the model to predicting outcome. the total sum of squares, denoted SS_T, measures the difference of the observed data from the mean of Y; the sum of squared residuals, denoted SS_R, measures the difference between the observed data and the regression line and the model sum of squares, denoted SS_M, represents the difference between the regression line and the mean of Y. If SS_M is large then the implication is that the model has made a large improvement in the prediction of the outcome, Y, when compared to just using the mean. The proportion improvement of the model can be determined thus:

$$R^2 = \frac{SS_M}{SS_T} \tag{4.4}$$

R^2 represents the proportion of variation in the outcome explained by the model and can be expressed as a percentage by multiplying R^2 by 100. SPSS provides another measure termed the F-Ratio that is based on the SS_M & SS_R values that is '…a measure of how much a model has improved the prediction of the outcome compared to the level of inaccuracy of the model [Field, 2000]. It is also possible to analyse the coefficients of the model. If a predictor variable is significantly contributing to the model then it should be significantly different from zero and this can be tested using a t-test. Finally, it is possible to assess residuals and model assumptions of normality using the 'P-P plot' provided by SPSS.

84 subjects were scanned on their volar and dorsal forearms using the Mexameter and SIAscope. The healthy adult subjects were recruited from staff and students from Addenbrooke's Hospital and the University of Birmingham. All categories of skin types were represented in the study. 2 sets of SIAgraphs were corrupted that were taken from the volar forearm as were 2 sets taken from the dorsal forearm and these were excluded from the statistical analysis.

4.2.2 Results

Linear regression analysis was performed on the dataset using SPSS for Windows v9.0 (SPSS Inc.) and Microsoft Excel 97 (Microsoft Corporation) and the results of this can be seen in table

4.1. In addition, the scatter plots with regression lines are shown in figures 4.1 to 4.4. From the table and it can be seen that the melanin data was encouraging. The size of the R^2 statistic indicates that a large amount of the variance can be explained by the model and the F-Ratio statistics indicate that the model greatly improves the prediction of the outcome compared to the its inaccuracy. However, the 'P-P plots' produced by the SPSS printout indicated that there may be a violation in the assumption that the data is normally distributed so it was decided to perform a Spearman's rho statistic (table 4.2). This is a non-parametric test to measure correlation between two variables and an in-depth description of this test can be found in Field (2000). The results of this test indicate that the Mexameter melanin value is highly correlated with the SIAscope melanin value and that a Spearman's rho of this magnitude is extremely unlikely to have occurred by chance.

Table 4.1 Regression Statistics for Mexameter and SIAscope Data

Dependent (Y)	Independent (X)	Equation	Coefficient Significance	R^2	F-Ratio	Significance F-Ratio
Mexameter Melanin (Inner Arm)	SIAscope Melanin (Inner Arm)	Y=942.7–1.961X	p < 0.0001	0.905	759.0	p < 0.0001
Mexameter Haemoglobin (Inner Arm)	SIAscope Haemoglobin (Inner Arm)	Y=568.0+0.694X	p < 0.0001	0.220	22.80	p < 0.0001
Mexameter Melanin (Outer Arm)	SIAscope Melanin (Outer Arm)	Y=885.0-1.706X	p < 0.0001	0.868	526.80	p < 0.0001
Mexameter Haemoglobin (Outer Arm)	SIAscope Haemoglobin (Outer Arm)	Y=594.0+0.171X	p = 0.172	0.023	1.898	p = 0.172

It may be noted that the Mexameter melanin values are negatively correlated with SIAscope melanin values (Figures 4.3 & 4.4) . This is purely an idiosyncrasy of the SIAscope data - the system inverts the pixel values so that melanin can be displayed as black on the monitor.

Figure 4.1: Scatter Plot of SIAscope v. Mexameter - Hb Inner Arm

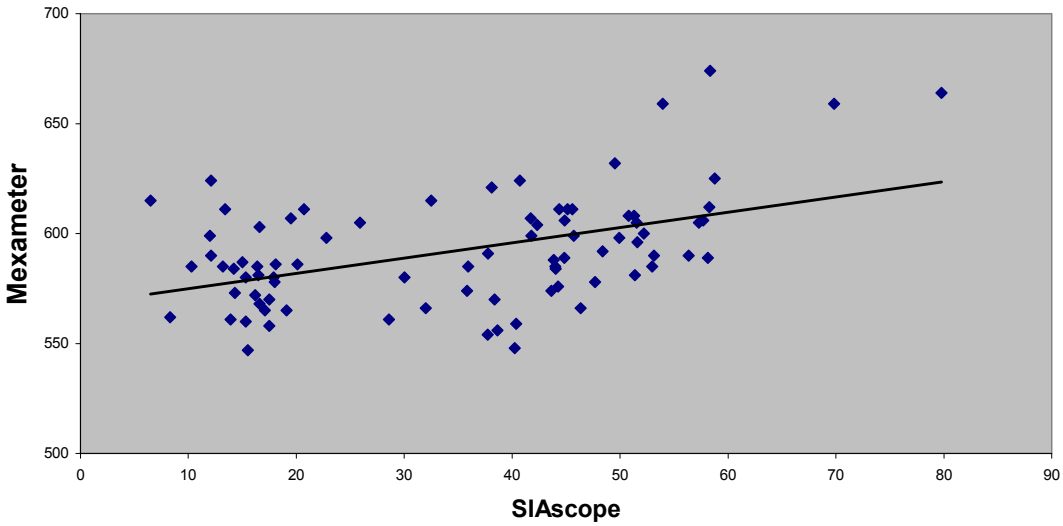


Figure 4.2: Scatter Plot of SIAscope v. Mexameter - Hb Outer Arm

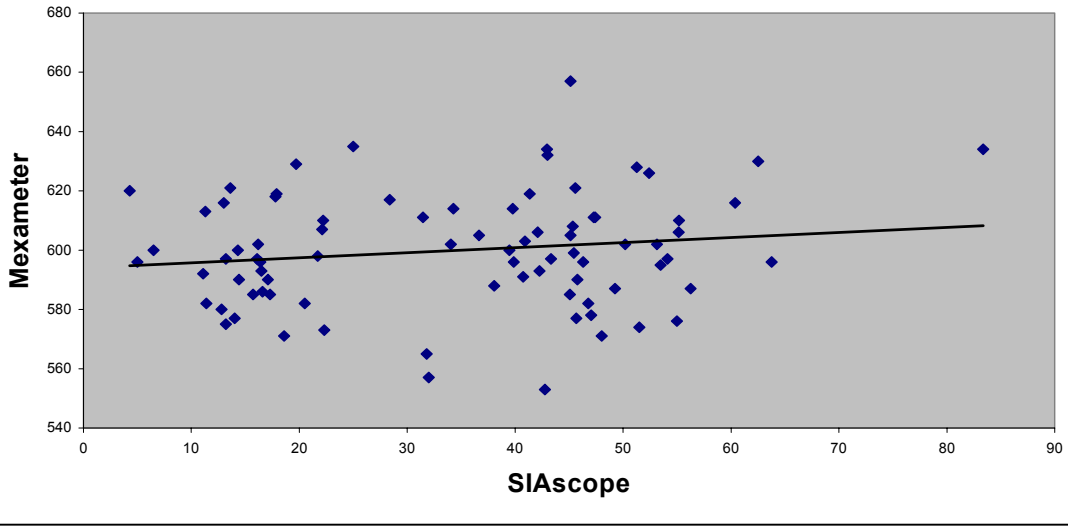


Figure 4.3: Scatter Plot of SIAscope v. Mexameter - Melanin Inner Arm

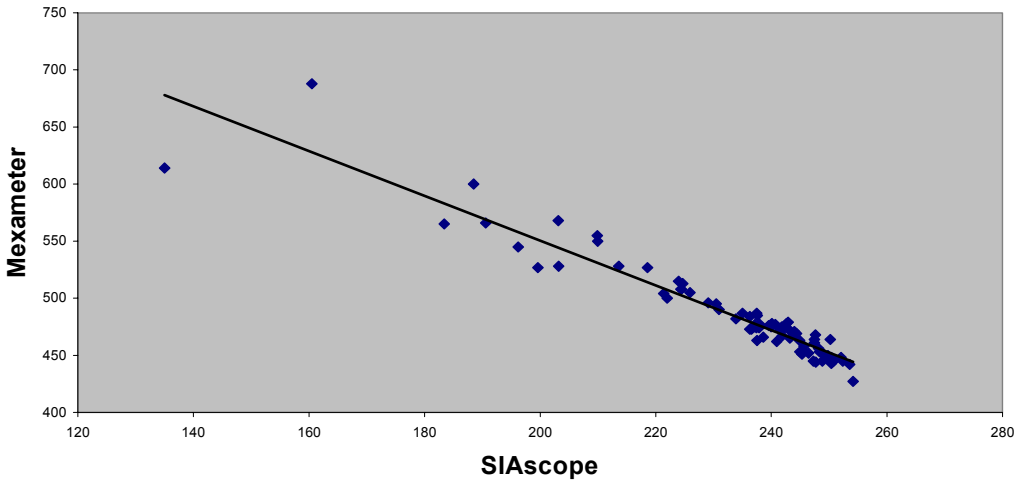


Figure 4.4: Scatter Plot of SIAscope v. Mexameter - Melanin Outer Arm

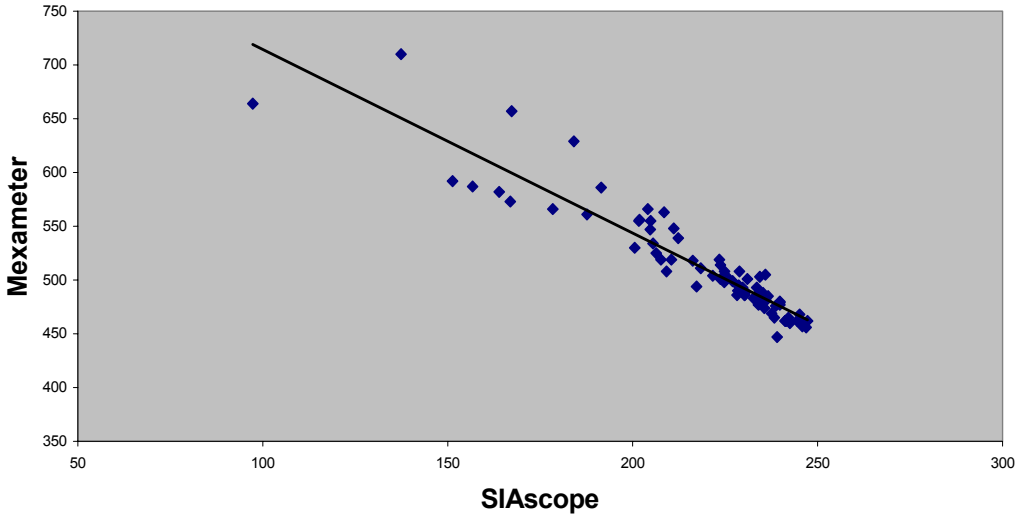


Table 4.2 Spearman's Rho Statistics for Mexameter | SIAscope Comparison

Dependent	Independent	Spearman's Rho	Significance
Mexameter Melanin (Inner Arm)	SIAscope (Inner Arm)	-0.949	$p < 0.0001$ (one tailed)
Mexameter Melanin (Inner Arm)	SIAscope (Inner Arm)	-0.956	$p < 0.0001$ (one tailed)

In contrast to the melanin data, the haemoglobin data were not so promising (table 4.1 and Figures 4.1 & 4.2). The R^2 values from the inner arm indicate moderate to poor correlation of the SIAscope haemoglobin value with the Mexameter and that a large amount of variance in the data remains to be explained. The F-Ratio was highly significant and the coefficient was significantly different from zero. These factors indicated that the model, although a poor one, significantly predicted the outcome better than the mean. However, the data from the outer arm was very disappointing in that neither the coefficient nor the F-Ratio achieved statistical significance and the R^2 value indicates that the model leaves nearly 98% of the variance of the data unexplained. It was felt that the explanation for this lay in the design of the experiment. At rest the blood flow through the skin of the arm is minimal and at this point both the SIAscope and the Mexameter are performing right at the bottom of their working ranges where sampling error is likely to occur due to the lack of sensitivity of the equipment. Increasing the sensitivity of the SIAscope is not possible as electrical 'noise' from the device would begin to be detected and produce further sampling error.

It was decided to perform another experiment that produced increased blood flow to the skin. Thus an experiment was devised that entailed measuring the haemoglobin concentration of the skin of individuals using the SIAscope whilst exercising vigorously on an exercise bike. The hypothesis was that the exercise would produce vasodilatation of the skin and the increase in haemoglobin would be detected in the SIAgraphs (figure 4.5). Eight subjects were recruited to the study and initial SIAscope and Mexameter readings were taken from the cheek before the exercising commenced. The subjects were then asked to exercise vigorously and SIAscope and Mexameter readings were taken from the same site at one-minute intervals until 9 minutes. Statistical analysis was performed using SPSS for Windows v9.0 and the tests performed were Wilcoxon's Signed Ranks test and the Sign test. The null hypotheses were that there would be no difference in haemoglobin concentration before and after as measured by the SIAscope and

Mexameter and it was also predicted that there would be a rise in haemoglobin concentration so that a one-tailed test would be appropriate. The results are shown in table 4.3

Table 4.3 Summary of Statistics for Comparison of Blood Measurements Before & After Exercise

Device	Time Interval (Mins)	Wilcoxon SR test Z value (N=8)	Wilcoxon SR test Significance (One-tailed)	Sign Test Exact Significance
SIAscope	9	-2.521	p = 0.006	p = 0.008
Mexameter	9	-2.240	p = 0.013	p = 0.07

From the table it can be seen that there was a significant difference in the haemoglobin concentration as measured by the SIAscope and the Mexameter after 9 minutes exercise both in the Wilcoxon Signed ranks test and the Sign test. In addition to these tests, linear regression analysis was performed on all the data from this second experiment. This was done to explore the possibility that the increase in haemoglobin concentration would reduce the experimental error between the two systems and thus improve the correlation between the two devices. The result is shown in table 4.4

Table 4.4 Regression Statistics for Exercise Experiment Comparing Mexameter with SIAscope

Dependent (Y)	Independent (X)	Equation	Coefficient Significance	R^2	F-Ratio	Significance F-Ratio
Mexameter Haemoglobin	SIAscope Haemoglobin	Y = 599.0 + 2.73X	p = 0.003	0.106	9.264	p= 0.003

The results show that, while the coefficient and the F-Ratio are significant, the R^2 value indicates that a large amount of variance of the data remains unexplained by the model.

4.2.3 Discussion

These experiments were designed to validate the SIAscope in measuring epidermal melanin and haemoglobin. However, when it came to designing these experiments it became increasingly apparent that they would be hampered by the lack of a decent ‘gold standard’ to correlate and compare with the SIAscope. Haemoglobin concentration is directly dependent on the blood supply of the skin and this is a dynamic and physiological parameter that is subject to rapid change and large fluctuations. It was therefore necessary to validate the SIAscope against a device that could reportedly measure these parameters *in vivo*. The Mexameter was chosen for these reasons.

The Mexameter and the SIAscope correlated almost perfectly in the measurement of epidermal melanin and it was concluded that the SIAscope had been validated for the measurement of epidermal melanin. This result was unsurprising as both devices are spectrophotometers that

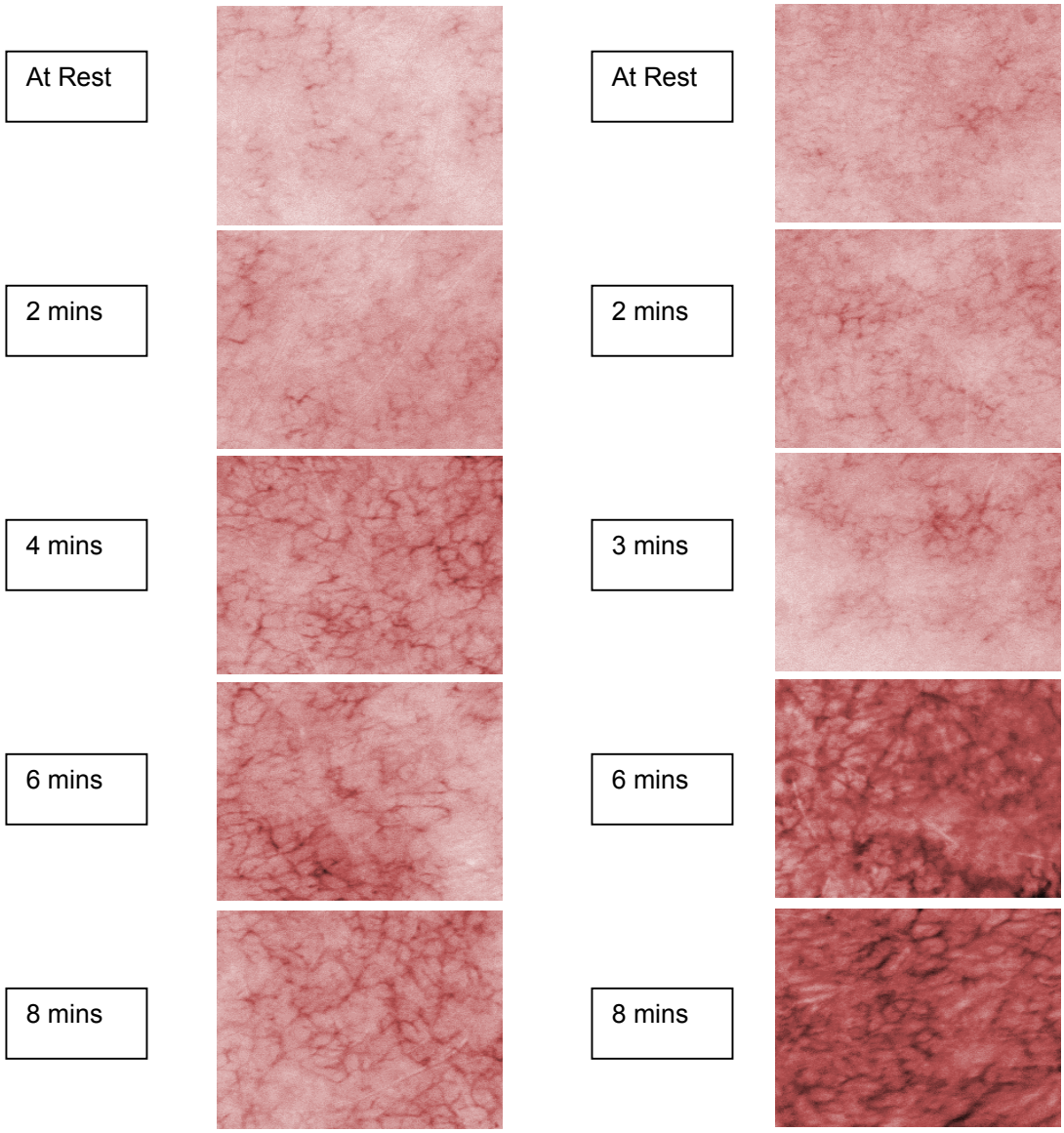


Figure 4.5:
Change in Blood SIAgraphs with time during heavy exercise from 2 volunteers

employ similar wavelengths to determine melanin concentration. In addition, as the melanin measured was only epidermal, the effect of the wavelength-specific remittance and absorption of light from the papillary dermis is not an issue and thus very similar results were expected. Correspondingly, if the Mexameter and SIAscope had been compared in the measurement of

epidermal & dermal melanin or just dermal melanin then it would be expected that the near-perfect correlation would no longer occur.

In contrast, the Mexameter and the SIAscope correlated very poorly in measurement of haemoglobin concentration. It was concluded that the Mexameter had not validated the SIAscope for the measurement of haemoglobin concentration. However, the Mexameter is a poor gold standard compared to the SIAscope – indeed, it could be said that the Mexameter does not come up to the standards of the SIAscope in determining haemoglobin concentration. From equation 4.2 it can be seen that the Mexameter does not use infrared light to determine the concentration of haemoglobin [Mexameter technical manual]. Cotton (1998) showed that the amount of light absorbed by haemoglobin is not only dependent on its concentration but also the depth that it is situated within the papillary dermis. In addition, the depth of the papillary dermis changes rapidly along a cross-section with the undulations of the dermo-epidermal junction caused by the papillary ridges. Any spectrophotometric system or device that fails to standardise the measurement of haemoglobin to the thickness of the papillary dermis is introducing significant measurement error.

A further source of discrepancy occurs as result of the differences in the size of the areas of skin measured by the two instruments. For instance, the SIAscope determined the haemoglobin concentration from an average of all values over an area of 225mm^3. In contrast, the Mexameter determines haemoglobin concentration over an area of just 5mm^3. Figure 4.5 demonstrates that, as the blood flow increases to the skin, the haemoglobin concentration does not increase uniformly; instead the distribution rather depends on the maximum diameter achieved by the cutaneous vessels as a result of vasodilatation. It is possible that if the Mexameter measured at a point in between these vessels then a spuriously low value would be recorded.

These two major sources of measurement error may explain why the SIAscope and the Mexameter correlated so poorly in the measurement of haemoglobin concentration.

The second experiment that measured the effect of exercise in the skin demonstrated a highly significant increase in haemoglobin concentration when determined by the SIAscope. The number of subjects in this experiment was small, and so any results have to be interpreted with some degree of caution. However, it was decided that the SIAscope had been validated for the measurement of haemoglobin for the purposes of continuing with the main part of this research in the analysis pigmented skin lesions. As it is discussed in section 3.2.1, it was expected that malignant melanomas would demonstrate erythema & inflammation and from this experiment it was concluded that the SIAscope would be sensitive enough to determine the presence of this.

Ideally, the experiment should be repeated to confirm this conclusion with a larger sample size or using a different erythrogenic method such as with a vasoactive cream or ultraviolet light.

4.3 Papillary Collagen

4.3.1 Introduction & Methods

Determining papillary collagen depth is the first and most important step of the SIAscope algorithm. Without obtaining this key information it is impossible to reliably determine the presence of dermal melanin or the concentrations of epidermal melanin and papillary haemoglobin. Thus, it was important to validate the system for papillary collagen and the 'gold standard' used was histology. However, there are several potential errors when comparing measurements obtained from histology with those from an *in vivo* system such as SIAscopy. The main source of error comes from the preparation of the specimen from biopsy to mounting on a slide. Skin that is still attached to the body is under a small degree of passive tension that results from the elastin and collagen fibres in the dermis. Once the specimen of skin is removed this tension is released, the skin crenates and the papillary dermis loses some of its form. In addition, the specimen is placed into formalin and/or alcohol as part of the fixation process and the dehydration results in further tissue shrinkage. Finally the tissue is mounted in a paraffin block and sliced using a microtome, some tissue separation and shredding may result from this. It is estimated that the specimen has usually shrunk by some 10-20% by the time it is finally mounted on a slide [Personal communication, Department of Histopathology, Addenbrooke's Hospital]. In addition to the shrinkage error, it is also important to be accurate in determining the region where the sample is scanned from as the composition of the papillary dermis can change dramatically from one anatomical site to another. For instance, the papillary ridges of glabrous skin are very pronounced whereas those on the face are virtually non-existent. Furthermore, the thickness of the papillary dermis dramatically decreases from the zygomatic region to the lower eyelid even though they are anatomically adjacent to each other.

Before deciding on the final experiments, it was necessary to decide what level of accuracy was required to validate the system given that there would be a large and unpredictable error incurred in the processing of skin from histology. It was expected that melanomas would show fibrosis with areas of absence of papillary collagen where invasive tumour occurred (section 1.3.2.2) and that benign naevi would display a regular pigment network that is a function of the contour of the dermo-epidermal junction (section 2.3.1.2). The features of fibrosis and 'collagen holes' were likely to be gross findings and the pigment network was estimated from skin surface microscopy images to require a lateral resolution of 40 micrometres. As a result, it was decided that a demonstration of resolution down to visualising papillary ridges would be adequate to validate the SIAscope for papillary collagen for the purposes of this thesis.

Patients were recruited for this experiment from the Department of Plastic Surgery at Addenbrooke's Hospital that were undergoing abdominoplasty or breast reduction surgery. As a result of the surgery there is redundant skin that is normally disposed and it was this skin that was scanned. Formal consent was obtained from the patients and the skin was then scanned and the region was marked to enable identification intraoperatively. Following removal of the redundant skin, the scanned region was excised, placed in formalin fixative and made into

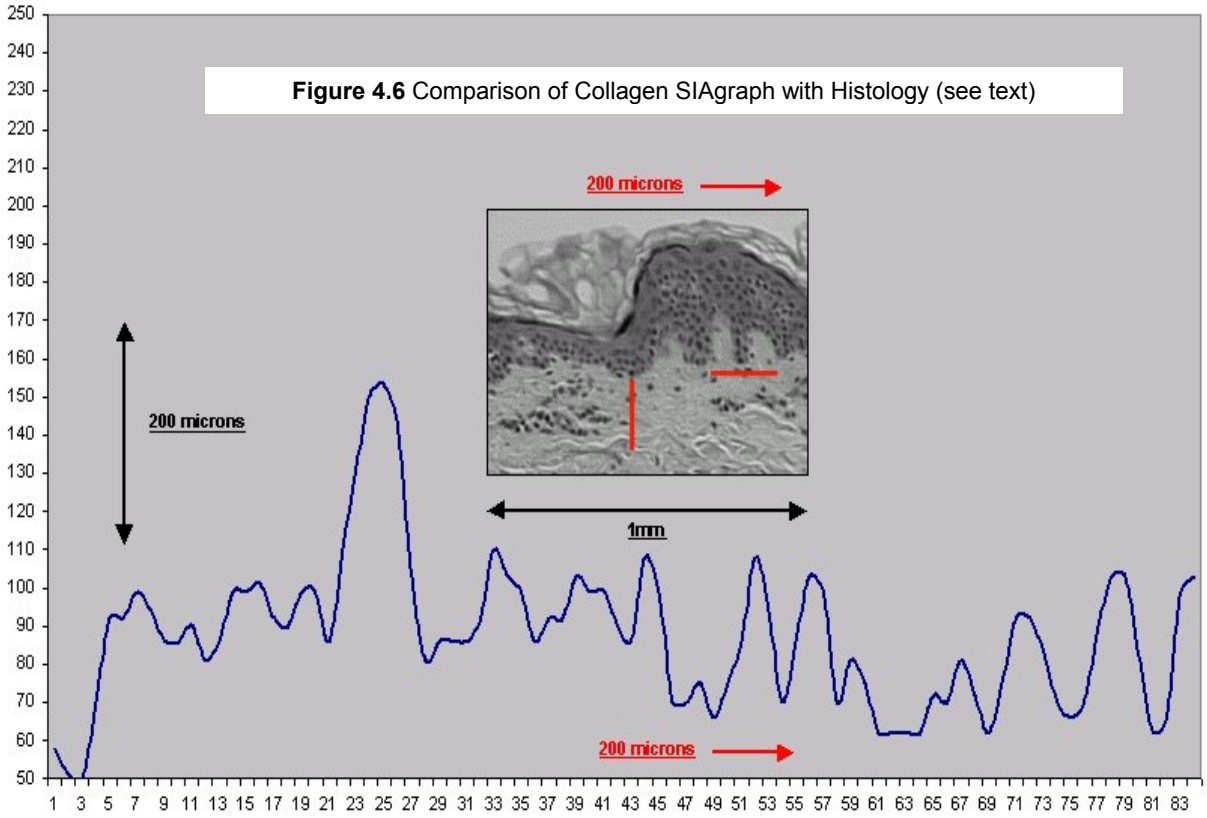


Figure 4.6 Comparison of Collagen SIAgraph with Histology (see text)

histology slides in the usual manner. Images of the histology slides were captured using a CCD camera (Panasonic P30) and an electronic frame grabber (Studio 400, Pinnacle Systems Inc, USA). In addition, the images were calibrated by imaging a reticule. Images obtained using the SIAscope were also calibrated by imaging a focus graticule.

4.3.2.Results

Five patients were involved in this study, generating five sets of data. The scanned images of the skin histology slides were analysed using Adobe Photoshop 5.0 (Adobe Systems Inc., USA) and Paintshop Pro 6.0 (JASC Software Inc., USA). It was calculated that in the collagen SIAgraphs 24 pixels represented 1 millimetre giving an absolute resolution of 42.0 micrometres. In addition, using the reticule, the histology images grabbed at 40-times

magnification displayed 1 millimetre every 290 pixels. Finally, a pixel brightness of 255 (maximum) represents 1 millimetre in depth of papillary dermis so that 200 microns, the expected average papillary dermal thickness [Anderson & Parrish, 1981], has a pixel value of 51. Using the software it was possible to sample horizontal strips of the SIAgraphs and determine the brightness value for each pixel. These values were plotted on a graph as shown in figure 4.6 that has a scanned image of the skin inset for comparison. Before discussing this figure, a few items must be pointed out first – whilst the inset image is to scale with respect to the graph horizontally, it is not to scale vertically; the y-axis of the graph starts at 50; the red bars in the scanned histology image represent 200 micrometres; finally the histology image and the cross-section are from the same $1cm^2$ region of skin but do not exactly match (it is therefore assumed that the skin in this region is uniform in structure). This graph is representative of all five datasets collected for this experiment. As can be seen in figure 4.6, large perturbations at the dermo-epidermal junction are detected well by the SIAscope. These are large papillary ridges with bases measuring greater than 100 microns across. However, the SIAscope cannot resolve smaller papillary ridges and two of them close together will be resolved as a single one, regardless of the maximal width of the individual ridges. The depth of the papillary dermis as measured by the SIAscope appeared to correlate well with the histology images. As can be seen in the figure, the papillary dermis varies from approximately 200 microns to 400 microns and this is reflected in the graph that ranges from 200 to 400 microns with a single peak that extends to 600 microns.

4.3.3 Discussion

This validation experiment offers no statistical analysis for formal objective assessment of the data. However, given the significant sampling error that occurs in processing the samples for histology it was felt that a simple analysis was still useful for the purpose of this thesis. The main aim of assessing the samples of normal skin was to ensure that there was no gross error within the system that could dramatically distort the results of subsequent feature analysis. In effect, the experiments were performed to calibrate the SIAscope for collagen. In section 2.6.3.3 it was discussed that SIAscopy is dependent on determining the papillary collagen depth to transform and standardise the skin colour-space to allow calibration and measurement of the parameters of haemoglobin and melanin and the accurate detection of dermal melanin. It is therefore not illogical to ask whether it reasonable to accept as true the other parameters returned by the SIAscopy if the accuracy of the collagen SIAgraph remains undetermined. This subject was discussed in great detail in Cotton's PhD. thesis (1998, Chapter 6).

An imaging system such as the SIAscope that attempts to measure collagen depth using near infrared primaries will be very accurate at small depths of collagen but becomes less accurate with larger depths [Cotton, 1998]. However, there is a non-linear relationship in the

transformation of the colour-space in the calculation of the values of haemoglobin and melanin. In effect, the potential for large errors is maximal at smaller depths whereas the potential for errors is minimal at the larger depths. This is a result of the movement of the sampled colour-points to the standardised colour-space being much larger with small depths than with larger depths [Cotton, 1998]. In summary, at smaller depths of collagen the system needs to be able to measure collagen very accurately which, fortunately it is able to do. Conversely, collagen measurements do not need to be so accurate at greater depths, which is fortunate because they aren't! In addition the system is also able to detect the presence of dermal melanin with great precision [Cotton, 1998]. If the collagen depth is under-estimated by the system then areas with dermal melanin will lie closer to the normal colour surface and there is a risk that they may not be detected. However, the only situation where collagen thickness is underestimated by the system is in the presence of very large quantities of dermal melanin [Cotton, 1998]. Happily, such large quantities produce a massive deviation off the normal colour-space such that non-detection is not an issue. Conversely, conditions that produce an over-estimation of collagen may cause the algorithm to consider a point has deviated off the normal colour-space when, in fact, it has not. This would produce false positives for dermal melanin. Clinical situations that could produce these conditions occur with lesions containing high quantities of keratin that is 'flaky' that results in large amounts of aberrant refraction and reflection. This situation can be resolved by using matching fluid such as oil or water in much the same way as performing skin surface microscopy.

It was anticipated that the useful feature that would be extracted from the collagen SIAgraphs would be collagen holes (section 3.6.2.2). In performing this experiment, it is accepted that collagen holes less than 80 microns in diameter would be difficult to detect. However, it was felt that the experiment had shown that the SIAscope was sufficiently calibrated for feature analysis in that it showed that the collagen SIAgraph measurements were within the same order of magnitude as the histology measurements. Other researchers are currently conducting experiments to accurately validate the SIAscope for collagen but currently this data is not available for this thesis.

4.4 Dermal Melanin

4.4.1 Introduction & Methods

It is expected that the presence of dermal melanin in a pigmented skin lesion will be a highly significant SIAscopic finding and so it is necessary to validate the system for this. Devising experiments to assess whether the SIAscope is detecting dermal melanin could provide two levels of evidence, namely direct and indirect evidence. Direct evidence that the system is faithfully detecting dermal melanin could come from comparing the SIAscope images with the

histopathology of the lesion. One approach could be to devise a prospective trial where lesions are excised, outside of any clinical need, from volunteers. The histopathology would be compared with the SIAscope images. This proposal was rejected on the grounds of ethics, expense and logistics. Therefore, the system has to be validated from the dataset that is collected for the main work of the thesis. According to guidelines outlined by Mooi & Krausz (1992a), in order to obtain tissue from a lesion for the purposes of research alone, the lesion should measure at most 8mm in maximal diameter and the section of study should not include the thickest part of the lesion or the nearest resection margin. Therefore, this limits any studies to the prepared slides used by the histopathologist to make a diagnosis. It is possible to use histochemical methods to stain for melanin that include Schmorl's or Masson-Fontana's stains or even Warthin-Starry's silver stain [Mooi & Krausz, 1992a]. False positives with these methods include lipofuscins and endocrine cells. In addition, it is possible to use immunohistochemical methods that stain for S-100 protein or HMB-45 and stains that use antibodies to type IV collagen in order to delineate the dermo-epidermal junction [Mooi & Krausz, 1992a]. However, these methods are expensive to perform and this becomes prohibitive when considering a dataset of roughly 600 lesions. Thus, whilst this method would be the ideal gold standard to assess and validate the SIAscope, the resources of the research group did not allow this.

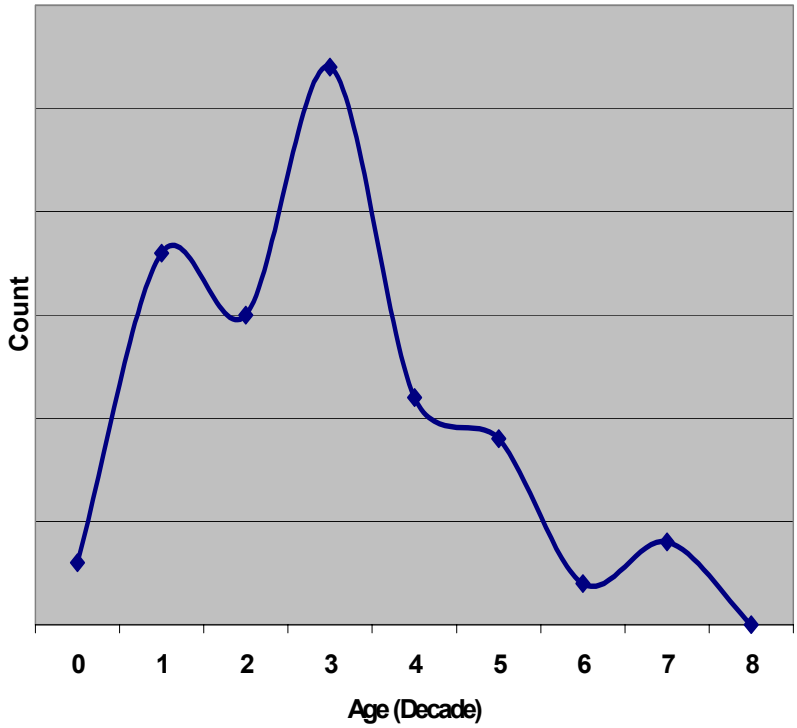
The alternatives for validating the SIAscope all rely on indirect evidence. One possibility that was considered would be to build a substitute model of human skin and insert melanin into the dermal layer. The model could consist of layers of gelatins or even porcine skin. The main problem with this method would be the necessity to rebuild the SIAscope algorithm so that it was tailored to the new model. This renders the method self-defeating as the experiment could easily be criticised for being applicable to the phantom model only and not to human skin. Another method that was considered would be to compare the histopathology reports with the SIAscope images. Thus when melanoma cells, melanophages, pigmentary incontinence or melanocytes were reported as present in the dermis, this could be cross-validated with the corresponding dermal melanin SIAgraph. However, these findings tend only to be infrequently reported, and only when they are of clinical significance so that there may be a bias to reporting them in malignant or dysplastic lesions and not in the benign lesions. In addition, the presence of dermal melanocytes does not necessarily correlate with the presence of dermal melanin – as dermal melanocytes can atrophy or 'mature' and no longer produce melanin (section 1.2.3). The final analysis considered was to correlate the presence of dermal melanin in the benign naevi, namely junctional, compound and intradermal and compare them with the age of the patients. It was hypothesised that the resulting graph would show a peak age between 20 and 40 that corresponds with the age-distribution of compound naevi in the adult population [Mooi & Krausz, 1992a]. In addition, chi-square statistics would be performed that

showed that the distribution of dermal melanin in the lesions was different between junctional, compound and intradermal naevi. This final method was the one that was chosen.

4.4.2 Results

Dermal melanin SIAgraphs of benign naevi were collated from both the main datasets collected from February 2000 until June 2001. In total 273 SIAgraphs were included in this dataset of which 89 indicated the presence of dermal melanin in the lesion. Figure 4.7 displays a graph of the distribution of lesions containing dermal melanin according to the age of the patient. It can

Figure 4.7: Distribution of Lesions Containing Dermal Melanin with Age



be seen in figure 4.7 that the number of lesions containing dermal melanin rises steeply to a peak in the first four decades and then gradually declines over the subsequent five decades. Table 4.5 indicates the distribution of the lesions with and without dermal melanin displayed in the Dermal Melanin SIAgraph according to the age of the patient. A χ^2 analysis using this table was performed. In order to meet the criterion that at least 80% of the cells must contain expected values greater than five, the totals of the first and second decades were combined in each group, as were the seventh to ninth decades [Bland, 1997]. The null hypothesis was that the presence or absence of dermal melanin in the Dermal Melanin SIAgraph of a benign naevus is independent of age. The test revealed that $\chi^2 = 14.433$ with five degrees of freedom and $p=0.013$. Accordingly, the null hypothesis was rejected. Table 4.6 displays the distribution

of lesions displaying dermal melanin on the Dermal Melanin SIAgraph according to the subtype of benign naevus. Similarly, a χ^2 analysis was performed based on this table. The null hypothesis was that the presence of dermal melanin in the Dermal Melanin SIAgraph of a benign naevus is independent of the subtype of that naevus. The test revealed that χ^2= 4.247 with two degrees of freedom and p=0.120. Thus, the null hypothesis could not be rejected on the basis of this data. To investigate the possibility of a trend in the decreasing presence of dermal melanin from the junctional and compound naevi to the intradermal naevi a χ^2 test for trend was performed on the data using the Mantel-Haenzsel trend test [Bland, 1997]. This analysis was performed using SPSS for Windows v.9.0. The null hypothesis was that there is no trend in the prevalence of dermal melanin in the Dermal Melanin SIAgraph with the subtype of benign naevus (from junctional to compound to intradermal naevus). The test reveals that χ^2 = 4.084 with one degree of freedom and p = 0.043. Accordingly the null hypothesis was rejected on the basis of this evidence.

Table 4.5 Distribution of Dermal Melanin in SIAgraphs According to Age

Age (Completed Decade)	**No Dermal Melanin Present**	**Dermal Melanin Present**
0	2	3
1	21	18
2	56	15
3	52	27
4	35	11
5	14	9
6	1	2
7	2	4
8	1	0
Total	**184**	**89**

Table 4.6 Distribution of Dermal Melanin in SIAgraphs According to Naevus Subtype

Naevus Subtype	**No Dermal Melanin Present**	**Dermal Melanin Present**
Junctional	11	8
Compound	122	66
Intradermal	51	15
Total	**184**	**89**

4.4.3 Discussion

The results of this study provide indirect evidence that the presence of dermal melanin in the Dermal Melanin SIAgraph directly correlates with the presence of dermal melanin in the lesion. As expected, plotting the distribution of lesions with dermal melanin with age on a graph demonstrated a peak incidence in the second to fourth decades. As was stated above, this correlates well with the findings of previous research that has shown that the intradermal melanocytes of compound naevi begin to atrophy and cease to produce melanin over time [Goovaerts & Buyssens, 1988; Mooi & Krausz, 1992a]. In addition, a χ^2 analysis of the data showed that there was enough evidence to suggest that the presence of dermal melanin in the SIAgraph is dependent on the age of the patient in benign naevi. Analysis of the distribution of dermal melanin in the SIAgraph with the subtype of naevus produced interesting results. The χ^2 test revealed that there was not enough evidence to reject the null hypothesis that the presence of dermal melanin is independent of the subtype of naevus. However, the χ^2 test for trend was significant at the 5% level. This situation is possible with both tests remaining valid. This is "… because the test for trend has greater power for detecting trends than the has the ordinary chi-squared test." [Bland, 1997]. One possible reason for the conflicting results of the distribution of dermal melanin across naevus subtypes is that junctional naevi often comprise of melanin-containing structures, such as melanophages, or display pigmentary incontinence in the papillary dermis. In contrast, it may be possible that the compound naevus consists of epidermal melanocytes that produce melanin and mature intradermal melanocytes that do not. Given that the proportion compound naevi that did not display dermal melanin was approximately twice as large as those that did, where one would expect the distribution to be evenly split, lends evidence to this theory.

It was decided that the analyses of the distribution of dermal melanin by age and naevus subtype provided enough evidence to consider the SIAscope validated for detecting dermal melanin and for the purposes of using this information in the feature-analysis studies of this research. In future it would be desirable to conduct studies that validate the system against histopathology using stains for dermal melanin.

4.5 Summary

This chapter was presented as a series of short papers outlining the experiments performed to validate the SIAscope for haemoglobin, melanin and collagen. In the first section, the Mexameter, an industry-standard spectrophotometric device used for determining tanning times and erythema by the cosmetics industry (amongst others) compared measurements of haemoglobin and epidermal melanin by the SIAscope. Measurements were taken from the skin on the forearms of healthy volunteers. When comparing melanin measurements, it was found

that the two devices correlated almost perfectly however, when comparing haemoglobin, the two devices showed little or no correlation. It was hypothesised that this was because the Mexameter did not take into account the depth of papillary collagen in its measurements and this would generate serious error [Cotton, 1998]. In other words, in the measurement of haemoglobin the SIAscope was hampered by not having a decent gold standard to compare it to. As a result an experiment was devised that would dramatically increase the blood flow to the skin to allow before & after comparison of the measurements of the SIAscope. This data showed significant difference in the haemoglobin content of the skin before and after the experiment. The SIAscope was considered validated for measuring haemoglobin and epidermal melanin.

Collagen SIAgraphs were validated by comparing images taken from normal skin with their histology. Experiments designed in this way suffer from errors that arise from processing tissue for histology. The skin usually crenates and loses up to 20% in volume in an unpredictable manner and it virtually impossible to correlate the exact point of imaging with the exact point of histology section. As a result, the experiments offered no statistical analysis except for calibration of the resolution of the SIAscope. In addition, the histology images were compared with plots of the cross-sectional values of the SIAgraphs (figure 4.6) indicating the system was performing adequately.

The SIAscope was validated for dermal melanin by comparing the distribution of the SIAgraphs of the benign acquired naevi. Previous research had shown that the intradermal melanocytes of compound naevi start to atrophy over time and cease to produce melanin [Goovaerts & Buyssens, 1988]. A comparison of the distribution of dermal melanin according to age was performed. This agreed with the supposition that the distribution of dermal melanin would rise to a peak in the first four decades and tail off thereafter, mirroring the distribution of compound naevi in the population [Mooi & Krausz, 1992a]. In addition, χ^2 testing revealed that the distribution of dermal melanin was dependent on the age of the patient ($p=0.013$). A comparison of the distribution of dermal melanin with subtype of naevi failed to prove that the distribution was dependent on subtype. However, a χ^2 test for trend showed that there is a trend in distribution from junctional to compound to intradermal naevus ($p=0.043$).

These experiments were simple in design and nature. This was mainly due to the fact that these were the first set of clinical studies to be conducted on the SIAscope and were performed in parallel with the main study of the thesis. Accordingly, any experimental design had to take into account the limits in time and resources of the research group. Direct evidence would come from complex laboratory-based experiments that were just not possible to undertake. As a result, a policy of calibration and detection of gross system error was adopted with the

additional provision of indirect evidence that the system is faithfully measuring the parameters that it purports to do. The results of these experiments demonstrate that, under these guidelines, the SIAscope was validated and calibrated. Ideally, the evidence of the laboratory-based studies would be desirable and, indeed some are currently underway [Personal communication, Dr. Cotton]. Nevertheless, these experiments provide encouraging results.