Gynecol Surg (2005) 2: 119–125
DOI 10.1007/s10397-005-0095-7

ORIGINAL ARTICLE

T. T. Carpenter · A. S. H. Kent

A new method of quantifying endometriosis using digital photography

Received: 19 December 2004 / Accepted: 18 February 2005 / Published online: 13 May 2005
© Springer-Verlag Berlin / Heidelberg 2005

Abstract The revised American Fertility Society scoring system for quantifying endometriosis is a relatively insensitive tool when assessing peritoneal endometriosis. We describe a new technique that can be used to quantify endometriosis which uses digital photography and a specifically designed computer analysis package to calculate lesion surface area. Using this we were able to demonstrate good intra-observer reproducibility, although inter-observer variability was relatively poor.

Keywords Endometriosis · Quantification

Introduction

Despite the fact that the revised American Fertility Society scoring system (rAFS) [1–3] is the standard technique for classifying endometriosis, it is rather limited in its discriminatory powers when dealing with peritoneal disease. By virtue of the fact that it was developed as a scoring system aimed to correlate with fertility, points are allocated heavily for disease affecting the fallopian tubes and ovaries with very little points being available for peritoneal disease. In fact, for peritoneal disease alone the highest possible score is six, which places it in the mild category. Thus if one is specifically interested in change in peritoneal disease the rAFS is a rather blunt tool.

In view if this, we undertook a pilot study to assess the feasibility and reproducibility of using digital imaging to assess the surface area of specific endometriotic lesions.

T. T. Carpenter (✉) · A. S. H. Kent
Dept. of Gynaecology, Royal Surrey County Hospital,
Egerton Road, Guildford, GU2 7XX, UK
E-mail: t.carpenter@doctors.org.uk

T. T. Carpenter
7 Pointout Road, Southampton, SO16 7DL, UK

Method

Subjects

This was a three-centre study with multicentre regional ethics committee approval. Six patients were recruited (two from each site). Inclusion criteria were patients who were due to undergo a laparoscopy for suspected or known endometriosis, were over 18 years old, had all gynaecological organs present and had given informed consent. There was no limitation on stage of disease.

Image capture

All patients underwent laparoscopy in the proliferative phase of the cycle. Laparoscopic entry was carried out in the usual way as per the practice of the gynaecologist, with particular care being taken when the uterus was instrumented not to be too vigorous with bimanual examination. On entry, the abdominopelvic cavity was inspected in the usual way, at which time an endometriotic ''index'' lesion was selected. This was a lesion that was anatomically relatively easy to photograph with no/minimal manipulation needed, next to which could be placed a needle. The operating camera system was then detached from the laparoscope and replaced by a Nikon COOLPIX 4500 digital camera with a special adaptor to allow attachment of the camera to the end of the laparoscope. The camera had 4.0 megapixel resolution.

Camera set-up

A straight 13 mm surgical needle (with 2.0 vicryl attached) was introduced via a second port and placed as close to the lesion as possible in the same plane. The light setting on the camera was set to incandescent with a light intensity of zero (range −3 to +3).

Using auto focus, supplemented by manual focus where necessary, the lesion was photographed close-up, ensuring that the whole lesion and needle was in view. The laparoscope was then pulled back, the camera refocused, and a wide-angle picture taken to allow the location of the lesion to be demonstrated.

The laparoscope and needle were then removed and the camera system detached. Following this, the whole process was then repeated to obtain a second set of images on the same lesion. The remainder of the laparoscopy was completed by the laser ablation of all visible endometriosis and closure of the abdominal port sites was by the usual practice of the operating gynaecologist.

Images captured on the digital flash card where then downloaded onto the hard drive of a Compaq Evo computer and copies of each pair (close-up and wide angle) of images were stored on individual compact discs.

Image analysis

The images were analysed independently by two gynaecologists using a specifically-prepared software package produced by Virtualscopics, Rochester, NY, USA. The 12 individual close-up images were presented in random order for each assessor to quantify. The surface area was calculated by first defining the needle for scale. The individual components of each lesion—red, black and white (as defined by the rAFS definition) [1]—were then circumscribed by the investigators using any combination of the various functions. Functions available were:

Filters:

Red, Green, Blue, Black, White

Delineators:

Live wire mode	Allows you to optimise the path drawn between successively user-defined points.
Shrink-wrap mode	Allows you to define a structure by roughly tracing the outside perimeter of a well-defined structure.
Region growth mode	Allows you to identify, with one mouse click, an entire well-defined structure.
3-D region growth mode	Similar to the Region Growth Mode, but the growth will proceed in three dimensions.
Geometrically constrained region growth (GEORG) mode	Allows you to define the shape of the geometric model. The model is used to smooth region boundaries and limit growth outside a structure of interest.
3-D GEORG mode	Operates on the same principle as GEORG, with the additional functionality of growth proceeding in three dimensions.
Add mode	Allows you to use free hand tracking to modify a currently finalized (red) contour by adding a new area.
Adjust mode	The Adjust Mode allows you to modify the active contour. Each time the left mouse button is clicked, the contour is forced to pass through the clicked point.
Continuous trace mode	Allows you to perform free hand tracing of structure boundaries.
Erase mode	Allows you to modify the currently finalized contour by using free hand tracing to delete the unwanted portion of the region.
Polygon mode	Allows you to manually trace structure boundaries by connecting points that the user made in the Image window. Straight lines are used to connect points defined by successive mouse clicks.
Rectangle mode	Allows you to define a rectangle.
Select mode	Allows you to convert a finalized (red) contour to an active (blue) contour.

When the investigator was happy with the defined area, the image was finalised and the computer calculated the surface area using the needle as a reference.

Statistical analysis

Intra- and inter-observer reproducibility was assessed by variance, coefficient of variation, and subjectively using plots.

Results

The total lesion area and the breakdown of red, black and white components as assessed by the two assessors are shown in Table 1.

The range in sizes of the index lesions selected for assessment is wide (0.5–69 mm^2). The index lesions for three out of the six subjects contained only red tissue and had no black or white component. These were also the three smallest lesions. The remaining three lesions had all three component areas, with the largest component being white scar tissue.

Plots of the lesion areas for the duplicate assessments made on each subject are presented by area type in Figs. 1, 2, 3 and 4.

The within- and between-subject variance components of the index lesion assessments are shown in Table 2 and the within-subject coefficient of variations (SD/mean) are shown in Table 3.

Table 1 Lesion surface areas as assessed by each assessor

Subject	Image	Lesion area (mm^2)											
		Total		Red		Black		White		Subject Mean		Centre Mean	
		Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2
1	A	0.54	0.48	0.54	0.48	0	0	0	0	0.54	0.55	1.32	3.04
1	B	0.53	0.62	0.53	0.62	0	0	0	0				
2	A	2.01	3.74	2.01	3.74	0	0	0	0	2.10	5.53		
2	B	2.18	7.32	2.18	7.32	0	0	0	0				
3	A	68.94	15.57	25.72	9.6	9.03	5.97	34.19	0	67.73	17.34	45.63	21.85
3	B	66.52	19.1	26.11	10.68	10.21	8.42	30.2	0				
4	A	19.13	17.94	1.75	0.91	0.25	0.6	17.13	16.43	23.54	26.37		
4	B	27.94	34.79	2.29	1.47	0.65	1.28	25	32.04				
5	A	14.38	11.4	14.38	7.3	0	0	0	4.1	14.29	11.63	22.29	19.42
5	B	14.19	11.85	14.19	8.91	0	0	0	2.94				
6	A	33.09	35.43	12.5	8.03	1.78	1.19	18.81	26.21	30.29	27.22		
6	B	27.49	19.01	9.68	6.65	1.56	1.55	16.25	10.81				

Fig. 1 Plots of total lesion surface area for each pair of lesions and for each assessor

Discussion

The ranges in lesion size and lesion components in these patients are quite wide, with some small lesions having only red components and other larger lesions being dominated by white areas.

To assess the reproducibility of a test there is no standard single "test" that can be applied. Instead one must use a combination of quantitative and qualitative assessments. In this experiment we used the coefficient of variation (CV) to quantitatively assess variability; however, this should be interpreted with great caution as the number of assessments is small and as such will be markedly influenced by any outlying values. As a general rule, a CV of 100 is used as the cut-off for an acceptable lack of variation, but the figure is obviously a continuum, with lower figures indicating less variability. The CVs obtained in this study for both assessors show good reproducibility for within-subject analysis, with all figures being below 100. Variability in all cases was less for assessor one than for assessor two, with the within-

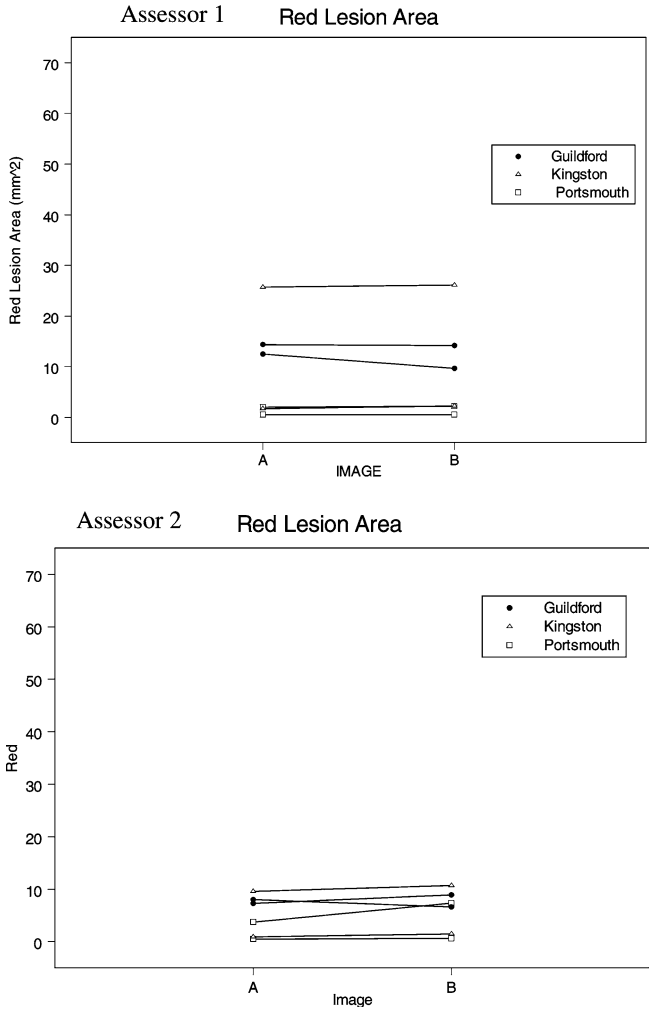

Fig. 2 Plots of red lesion surface area for each pair of lesions and for each assessor

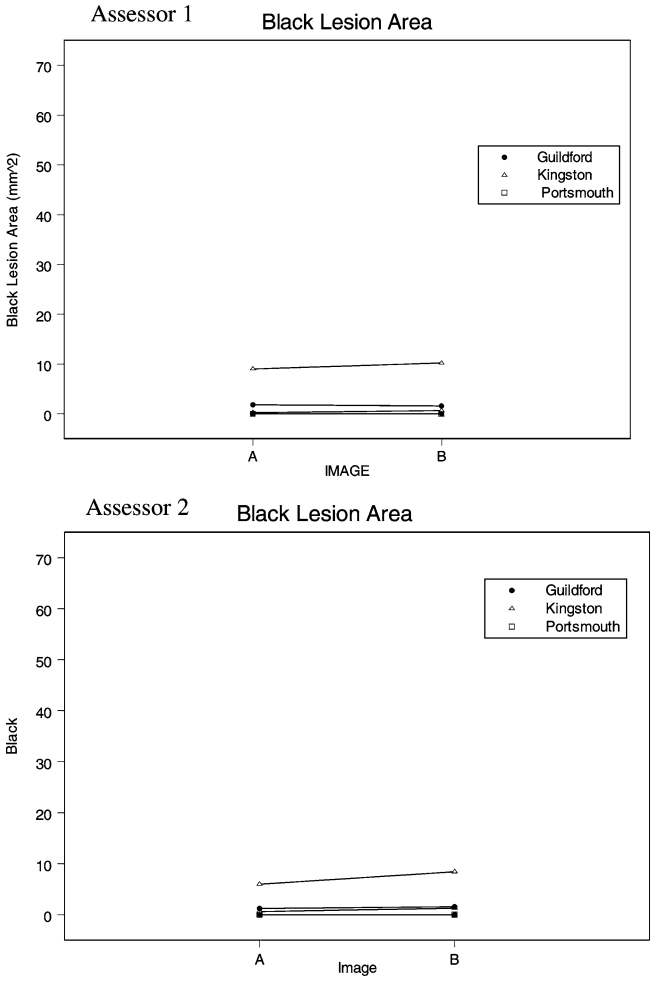

Fig. 3 Plots of black lesion surface area for each pair of lesions and for each assessor

subject variation contributing between only 0.7 and 3.6% of the total variance of assessor 1 compared to a within-subject variance for assessor 2 of between 6.7 and 34%. However, it is important to note that the percentage variance figures given in Table 2 are percentages, so the figures are proportions. Therefore, if the between-subject variance component is high, as in the case of assessor 1, then proportionately, and thus as a percentage, the within-subject variance will be low. This can therefore give an artificial impression of a smaller within-subject variance than is actually the case. Looking at the true figures for variance, it is still clear that assessor 1 gave consistently less within-subject variability than assessor 2, although to a lesser extent than is apparent from the percentage values.

Upon reviewing the techniques used by both assessors, assessor 1 used significantly more manual tracing of the regions via the live wire mode than user 2, who predominantly used the more automated functions of regional growth and geometrically constrained regional growth modes. This would imply that, although more time-consuming, semi-automated manual drawing of the lesions is a more reproducible technique.

In both cases the most reproducible component of analysis was the red area. This is encouraging, as red areas are considered most active [4], and thus if one were to test any new treatment for active endometriosis, one would expect these areas to respond first and possibly to a greater extent.

The total CVs for the two assessors were clearly far less reproducible. As a function of both between-subject variance and within-subject variance, this would be expected due to the wide ranges in lesion sizes and compositions. When assessing the efficacy of any treatment (or indeed a simple change in disease) one is interested in the changes in individual lesions or components of individual lesions, not the changes in total areas of lesions in a combined set of patients. As such this variation is unimportant.

As explained earlier, assessment of reproducibility is enhanced by the combination of both quantitative and qualitative methods. Thus the plots of lesion size are equally important when drawing conclusions on

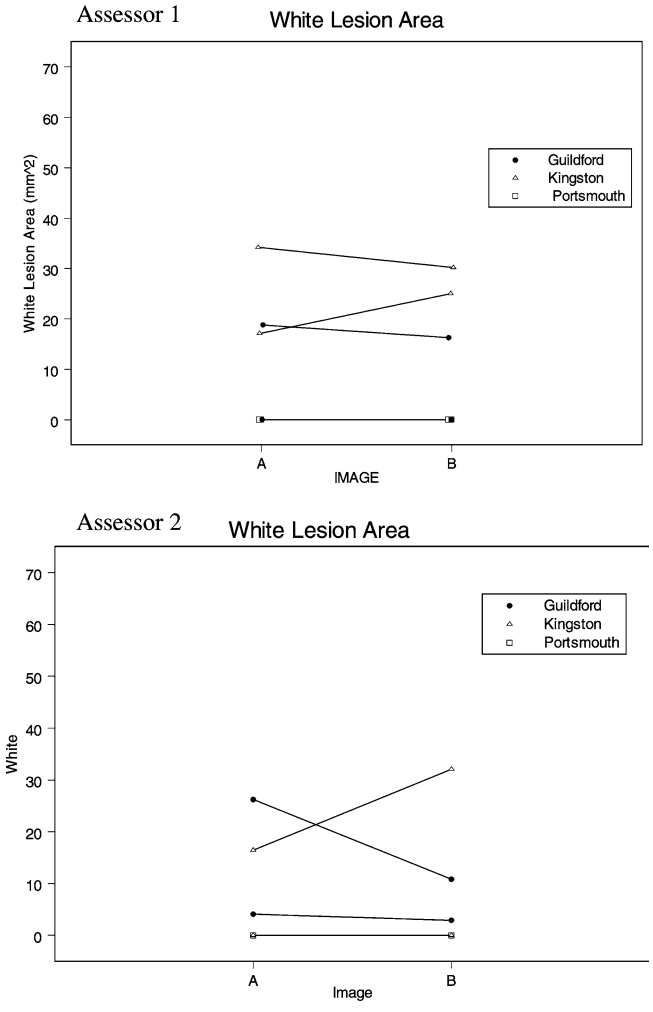

Fig. 4 Plots of white lesion surface area for each pair of lesions and for each assessor

reproducibility and they also allow us to assess intra- and inter-observer variability. From Tables 1 to 3 and Figs. 1, 2, 3 and 4, the intra-observer reproducibility of the red and black areas appears very good in both observers. Note, however, that assessor 1 found no black areas in three of the subjects and assessor 2 found no black areas in two subjects, thereby making the plots of the black areas for the six subjects look better than perhaps is truly the case. Red areas, however, were present in all subjects, and the reproducibility is good throughout. The total lesion areas and white lesion areas do seem to show greater variability, with the greatest variability being apparent in the larger lesions. The reduced reproducibilities of these areas compared to the red and black areas are probably related to the fact that the red and black areas have more obvious borders and thus confines. We are looking at lesions on a background of peritoneum which itself has a white/greyish appearance. Identifying a border between a white area of endometriosis and normal peritoneum is therefore considerably harder, and hence more prone to error, than defining a border between a red or black area and

Table 2 Variance of lesion surface areas for each assessor

Lesion area	Between-subject component variance		Within-subject component variance		Total variance		Proportion of total variance from between-subject variance (%)		Proportion of total variance from within-subject variance (%)	
	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2
Total	609.58	94.6	9.57	48.3	619.15	142.9	98.5	66	2.5	34
Red	96.82	14.17	0.71	1.57	97.53	15.74	99.3	90	0.7	10
Black	14.45	7.62	0.13	0.55	14.58	8.17	99.1	93.3	0.9	6.7
White	186.95	97.03	7.03	40.18	193.98	137.21	96.4	70.7	3.6	29.3

Table 3 Coefficients of variation for each assessor

Lesion area	Overall mean		CV (Total)		CV (Within-Subject)	
	Assessor 1	Assessor 2	Assessor 1	Assessor 2	Assessor 1	Assessor 2
Total	23.08	14.77	107.8	80.93	13.4	47.1
Red	9.32	5.48	106.0	72.4	9.0	22.9
Black	1.96	1.58	194.8	180.9	18.4	46.9
White	11.80	7.71	118.0	151.9	22.5	82.2

normal peritoneum. The greater variability seen in the larger lesions is most likely to be as a consequence of size. The potential for error is going to increase with lesion size since the border to be defined is longer. In addition, we show actual lesion size in the plots rather than relative differences between plots. Thus, a 20% difference in actual surface area between plots in a small lesion will be markedly smaller than a 20% difference in a larger lesion. In reviewing plots of the two assessments, the larger lesion will have a much steeper slope between the two plots than the smaller lesion, despite the relative difference in the two assessments being the same.

Assessment of inter-observer variability is best made by viewing Figs. 1, 2, 3 and 4. The more horizontal the line, the lower the intra observer variability, and the closer each pair of lines lie to each other, the lower the inter observer variability. From all plots it is clear that the reproducibility of analysis is significantly worse between assessors. The differences between the assessors when assessing the black areas does not seem great; however, as previously mentioned this is because no black areas were present in two of the subjects (as assessed by both assessors), and with the exception of subject 3 the black areas in the other subjects were very small. The lesion from subject 3 showed wide variation in all areas of assessment between the two assessors. On review, this lesion is a very complex lesion, with a potentially large chance of error.

There was no consistent difference between the assessors, although it is interesting to note that whilst the red area was most reproducible in within-observer analysis, it was in fact the area that showed most variation between observers. Thus it would appear that, although borders of these areas are easier than other areas to elucidate, the subjective decision as to whether an area is red, black or white still remains to be made by the assessor, and this appears to show more variability.

When assessing reproducibility one must be aware of the components contributing to the variability. The tables and figures concentrate on the variability within and between the two assessors. In fact the pairs of lesions analysed in each case are not the same image of the lesion but two different images of the same lesion taken at different times (albeit the same operation). Thus the within-subject variation is not only a function of the variability of the assessor but also the variability of the image taken. Whilst efforts have been made to minimise the variability between the images, inevitably the photographs will not be taken at exactly the same angle, the needle will not be in exactly the same position for each photograph, and other variables will be slightly different between them too. It is important to be aware of these differences, as they are likely to be minimised in this study by the fact that the images were captured during the same operation a short duration apart. In studies assessing the efficacy of a treatment over time, the images will be captured during separate operations some time apart, almost certainly increasing the variability.

The aim of developing new analysis techniques is to facilitate the detection of clinically-significant effects of a treatment on a disease, which in this case means that we would like to detect a clinically-significant difference in endometriotic lesions. In endometriosis, whilst one can make assumptions as to what would be a clinically-significant reduction in pain for example, because of the lack of correlation of symptoms with disease, it is not possible to logically select a value for clinically-significant change (with the exception of total disease irradiation) in lesion surface area. This technique is therefore most useful as a research tool. For ethical and economic reasons it is not possible to run long, large-scale trials of new treatments without some indication of the efficacy to begin with. Research tools such as this will allow investigators to detect changes, which may be relatively small, over a short period of time, which may then be used to justify a larger, more pragmatic study of a particular treatment.

What is important, however, is to be aware of the ability, or limitations, of a test used to detect a difference. A test that has a relatively high variability is going to be unable to detect small differences between two groups because the difference will be "masked" by the background intrinsic variability. From the results obtained using this technique, assessor 1 should be able to detect a within-subject variance of more than 3.6%, and assessor 2 a within-subject variance of more than 34%. These figures should be kept in mind when interpreting the results. Because of the relatively small number of lesions used in this study and the differences between each pair of lesions examined, it was not possible to assess whether there was a "learning curve" with this technique.

One criticism of this technique is that it only measures the diseased surface and takes no account of depth of invasion and volume of disease. This would be particularly applicable to patients with relatively advanced disease or those in areas such as the uterosacral ligaments, which often have little visible disease at the

surface but have deep deposits within. At present, the most likely route to quantifying this sort of disease would be via other means of imaging, such as magnetic resonance imaging. While work has been undertaken in this area, a reproducible technique has not been found to date [5].

Conclusions

This technique demonstrates acceptable intra-observer variability for both assessors; however, there is significant operator dependence for reproducibility. The intra-observer reproducibility is relatively poor. This technique allows us to estimate an observers variability, thereby highlighting what might be considered a clinically-significant change when power calculations are performed for future studies.

Acknowledgments We would like to acknowledge Pfizer UK for provision of the digital equipment and statistical analysis. We would also like to acknowledge Mr. S. Ewen and Mr. A. Pooley who undertook the digital photography at two of the sites.

References

1. American Society for Reproductive Medicine (1997) Revised American Society for Reproductive Medicine classification of endometriosis. Fertil Steril 67(5):817–821
2. American Fertility Society (1985) Revised American Fertility Society classification of endometriosis. Fertil Steril 43(3):351–352
3. American Fertility Society (1979) American Fertility Society for Classification of endometriosis. Fertil Steril 32(6):633–634
4. Khan KN et al (2004) Higher activity by opaque endometriotic lesions than nonopaque lesions. Acta Obstet Gynecol Scand 83(4):375–382
5. Kinkel K et al (1999) Magnetic resonance imaging characteristics of deep endometriosis. Hum Reprod 14(4):1080–1086