Observations on the accuracy of point counting including a description of a new graticule

BERNARD LENNOX

From the University Department of Pathology, Western Infirmary, Glasgow

SYNOPSIS The accuracy of point counting as a histometric technique depends very much on the number of points counted, and the number necessary varies greatly with both the level of accuracy required and the proportion the relevant tissue occupies in the section. It is shown that the number of 'hits' is a more useful measure than the total number of points counted, and a simple rule derived for determining the number of hits necessary to achieve any necessary degree of accuracy. A method of progressive estimation of accuracy when attempting to measure the proportion of a tissue that is unevenly distributed is also described.

The point-counting graticules commonly used, with 25 points, are inadequate with tissue proportions below 10%. A new graticule design, with 100 points optimally distributed, and usable as a 25-point array if required, is illustrated.

Of quantitative techniques in histopathology, one of the simplest and most generally useful is the point-counting method of determining relative areas in a section (Dunnill, 1968), which may be used, for instance, to determine the volume of any given tissue or structure in an organ. Observation of the manner of presenting results by various authors using this method suggests that most users have very little idea of the degree of accuracy and reproducibility to be expected from it. We have been using it extensively in studies on the testis (Dykes, 1969; Ahmad, Lennox, and Mack, 1969; Lennox, Ahmad, and Mack, 1970; Ahmad, Dykes, Ferguson-Smith, Lennox, and Mack, 1971) and had often to determine whether differences between testes (in volumes of Leydig cells particularly) were significant. From this difficulty arose the question, How many points are required to be counted to achieve a given degree of accuracy? This paper attempts to answer this question, and offers some suggestions on the practical implications of the answer, including a new graticule design.

These comments must not be regarded as a criticism of point counting in general. The level of accuracy obtainable is probably better than with any other available technique, and the labour certainly less. But one must not allow oneself to be misled by the spurious appearance of accuracy of results based on counts of a few score of points.

Assuming Random Distribution

In many situations (as for instance Leydig cells in a normal testis) the tissue being studied is distributed sufficiently evenly through the organ for the variations from field to field to be regarded as random. The accuracy then approximates closely to that expected on theoretical analysis. This is the ideal situation in which results are the best attainable.

Point counting assumes that the chance that a randomly positioned point of an eyepiece graticule will fall in a given tissue corresponds exactly to the proportion of the area of the section occupied by the tissue. This is statistically the same thing as the balls-in-a-bag problem. The chance of drawing a black ball from a bag of mixed black and white balls corresponds exactly to the proportion of black balls in the mixture.

If, then, a sample of n balls is drawn from a bag containing $p\%$ of black balls and $q\%$ of white (p and q of course totalling 100), the proportion of black balls in the sample will give an estimate of the true proportion ($p\%$) which will be liable to considerable error if the sample is small but improve as it enlarges. The size of the error is given in terms of the well known formula for the standard deviation (s), which applies equally to a 'sample' of points in point counting.

$$s = \sqrt{\frac{p\ q}{n}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(1)$$

The size of the standard deviation is, however, of little significance except in relation to the size of the proportion being measured: thus an SD of 5% indicates a very useful estimate of a proportion of 50%, but a meaningless estimate of one of 1%. We can use the ratio (standard deviation/best estimate of proportion) as a measure of the reliability of the estimate. Calling this ratio R, we can derive from equation 1 a new one for the number of points to be counted to give any required level of R.

$$n = \frac{q}{R^2 p} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots(2)$$

Figure 1 shows the result of applying this formula for two representative levels of R. It is at once
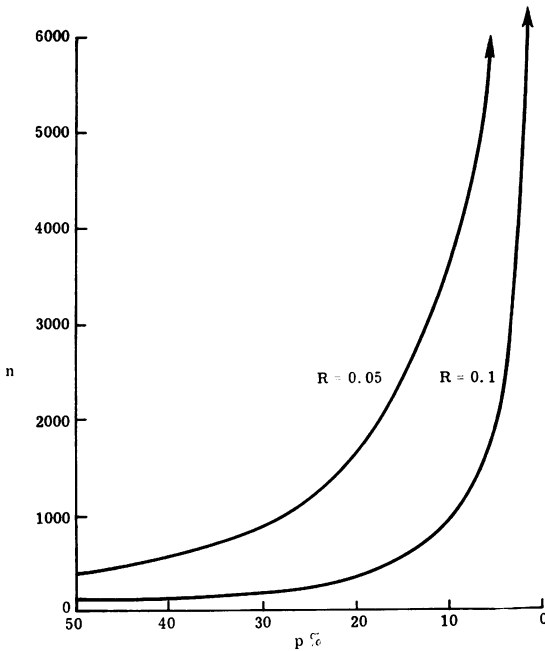

Fig 1 *Graph showing the effect of variation in the proportion (p) of the area of the section occupied by the tissue under study on the number (n) of points required to be counted to produce an estimate, with two representative levels of accuracy, R. (R is produced by dividing the standard deviation of the estimate by the proportion.)*

evident that if the relevant tissue is present in low concentration the number of points to be counted becomes extremely high. (If the proportion of tissue was as low as 0.1%, to achieve an R of 1 in 20 would require 400 000 points.)

If this were the whole truth, point counting for low proportions would be impracticable. If, however, we consider not the total points, but only the positive points or 'hits', the formula becomes

$$n \text{ (hits)} = \frac{q}{100\,R^2} \quad \dots\dots\dots\dots\dots\dots(3)$$

The figures are now much more manageable (fig 2). The effect of the proportion is much less, and there is an upper limit of hits above which one need not go, no matter how small the proportion. This limit is simply $1/R^2$. It follows that it is exceedingly expensive to achieve extreme accuracy, and levels of R much below 1/20 are not practicable in most cases.

A SIMPLE RULE OF THUMB
The real value of the above demonstration is this. The relation of total points to accuracy is so dependent on proportion that it is not possible to determine the number of points before the proportion is known fairly accurately, and this cannot be known till the count has been done. Using hits, the proportion can either be altogether neglected, or only a very rough estimate of it be used. The following simple rule emerges. For any desired degree of accuracy, expressed in terms of R, square the reciprocal of R, and go on counting until reaching that number of hits. If the proportion expected is over 10%, the number of hits can safely be reduced by the amount indicated by fig 2 and formula 3, but never by more than 50%.
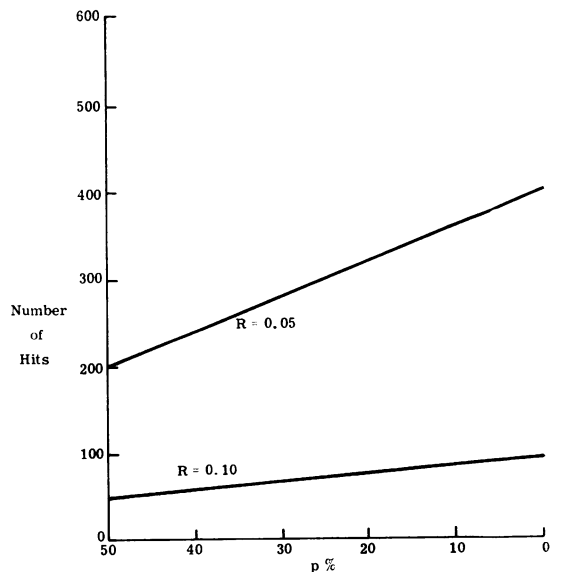

Fig 2 *As fig 1, using positive points ('hits') instead of total points, showing the much smaller influence of proportion: note also the linear relation and the defined maximum value for each value of R.*

In deciding on the necessary level of R, remember that (except with small numbers of observations) two estimates are usually significantly different if the difference exceeds the sum of their individual standard deviations. Therefore if a figure can be given for the smallest difference which it is hoped to be able to detect, half this determines the highest acceptable standard deviation. For example, if the operator will be satisfied to distinguish reliably a proportion of 9% for one of 11%, there is a difference of 2%. Half this is 1%: R is the SD divided by the proportion, and so in this case about 0·1. Squaring the reciprocal of this gives 100. For safety, a goal of 120 hits would be reasonable.

The use of hits rather than total points has consequences for the technique of counting which will be considered later under graticule design.

Non-random Distribution

If the tissue being studied is irregularly distributed in the organ (as Leydig cells in a scarred testis) the estimate of proportion cannot be as good as that made on a regular distribution, or, in other terms, to achieve the same degree of accuracy it is necessary to count even more points. With major degrees of irregularity, no pretence of accurate estimation should be made. If it is attempted in intermediate cases, it is particularly important to ensure that the fields examined are selected either entirely at random, or according to some rigidly applied formula that eliminates any kind of bias in the choice. It still remains necessary to find some way of determining whether enough points have been counted to give a result which, in spite of the irregular distribution, has an adequate degree of reliability. The following method has worked reasonably well in my hands.

If the number of hits per field averages over 20, single fields suffice, but if it is lower than this a number of fields must be grouped together to give an average total of hits per group of over 20. Counts are then recorded as totals for each such group. Having counted about 10 such groups, there is enough information for a preliminary estimate of proportion and standard deviation. (A rapid method of estimation of the SD of the individual group scores is described below.) Dividing this SD by the square root of the number of groups counted gives the SD of the mean for all the groups so far counted. From the estimate of proportion now available, and knowledge of the level of R required, it is possible to tell what level of SD is acceptable and how this compares with the SD so far available. And since it is known that the observed SD ought to fall in inverse proportion to the square root of the number

of observations, a forecast can be made of roughly how many groups must be counted to attain the required accuracy. A decision may often be taken at once whether to continue or to abandon the exercise. In most cases one will continue tentatively, repeating the calculations at intervals. If the SD of the mean continues to fall reasonably regularly in accordance with expectations and the proportion remains steady, the exercise is worth while even though the degree of accuracy attainable with the number of points one is prepared to count may not be high: but if these parameters continue to behave irregularly, it must be recognized that nothing will be gained by continuing.

ESTIMATING THE SD
The above method depends on rapid and easy estimation of the SD. A programmable desk calculator may be used with a program that allows progressive addition of values, since otherwise it will be necessary to start from scratch for each estimate. A rough but adequate estimate of the SD can, however, be obtained very quickly as follows, remembering that in a normal distribution only one value in 20 will lie outside a span of 4 SDs, ie, 2 SDs on either side of the mean. For 20 values or less the SD may be taken as a quarter of the range. For more than 20 figures, delete one figure in 20, choosing the most divergent, and take a quarter of the range of those remaining. This kind of estimate of the SD is often adequate in situations other than the present: accurate calculations can be reserved for cases with significance on the borderline. For large samples, the difference between the means of the top and bottom 5% is the best measure of 'range' for this purpose (Quenouille, 1959).

Practical Counting and Graticule Design

Though in principle point counting assumes the random movement of a single reference point, and it is quite possible to work with any casual speck of dust that can be brought into focus with the specimen (eg, Dykes, 1969), in practice an eyepiece graticule presenting an array of points on a glass disc is normally used. The necessary characteristics of this are as follows:

(*a*) It must be easy to bring all the points into focus with the specimen, and it is a great advantage if the field is flat enough to allow all points to be in focus at the same time. Though an ordinary eyepiece can often be found that matches a given graticule, in practice it is best to use a special eyepiece with independent focusing which can accept a variety of graticules.

(*b*) The points must be fine enough for accurate

localization, but easy to identify. Crossed lines are better than round spots.

(*c*) The points should be arranged in some regular manner, to ensure that all of them are inspected. The Chalkley design, with 25 coarse spots irregularly arranged, fails on both (*b*) and (*c*).

(*d*) The number of points should be related to the type of counting being done. The Leitz 'integrating eyepiece', with 25 points arranged on parallel chords of a circle, is an admirable design, but 25 points is too few for many purposes. Moreover, this expensive eyepiece cannot be used for other purposes, eg, with other graticules.

THE IDEAL NUMBER OF POINTS IN A GRATICULE
If the tissue being counted is abundant, the natural process of point counting involves scanning each of the points in a field in turn, and counting the hits as they turn up. For this, 25 points is convenient: with more, it is easy to lose count of the hits in a field. Moreover, this ensures that a reasonable number of distinct fields are examined, and it is always wise to do this if possible. But with a less abundant tissue— under 10% usually, and certainly under 5%—it is easier and quicker first to identify the positive areas in the field, if any, and then to count the number of points falling on them. A graticule with as many points as possible will be found very much more satisfactory for this method of working. An additional consideration is the fact that with a small biopsy, the number of possible fields may be limited (a graticule may be rotated to obtain multiple counts from one field but this is not very satisfactory).

The number 100 is very convenient for calculation, and it seems unlikely that very many more points could be accommodated in an eyepiece graticule without excessive crowding. A suitable eyepiece graticule is produced by Messrs Graticules Ltd[1], under the code name of G14: it has 100 crosses in a 10 × 10 square, and was designed by Dr D. W. Humphrey of the Department of Geology of the University of Sheffield (personal communication). It proved very satisfactory, and was used in most of our work on the testis already reported. With this device, counting of low-proportion tissues such as normal Leydig cells is easy. The operator counts with a hand counter in each hand, scoring fields on one and hits on the other. When the appropriate number of hits has been recorded, division of that figure by the number on the fields counter gives the percentage direct. (In routine testis work, 200 hits has been the usual level accepted, with 400 or more in cases of special interest.) With any less convenient system, counting the number of points required for a low-frequency tissue is very laborious.

[1]Messrs Graticules Ltd, Sovereign Way, Tonbridge, England.

A graticule of 100 points seems ideal for counts in the 1-10% area. For proportions much below 1% a much larger number of points would be better. It is, however, unlikely that any eyepiece graticule could cope with this: a projection system of some kind would probably be necessary.

A NEW GRATICULE
The simple 10 × 10 graticule is not altogether satisfactory because it is square, and hence wastes over a third of the area of the circular field. Experiment with possible alternatives showed that a very simple design, based on a 9 × 9 square array with four or five additional points on each side, made a much more efficient use of a circular space (fig 3).
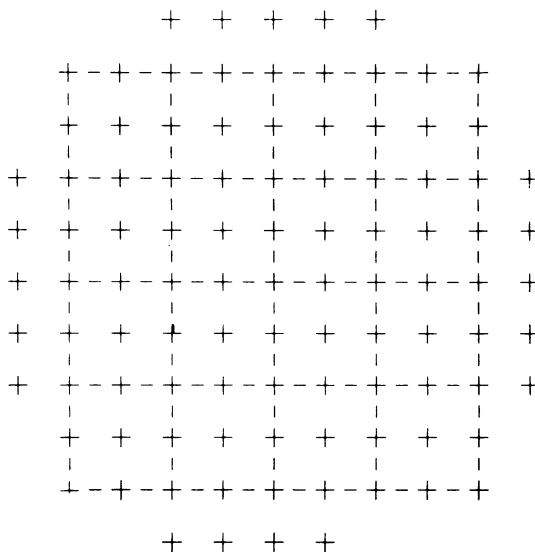

Fig 3 *The design of a 100-point graticule recommended particularly for counts at proportions below 10%. Note that the points at the corners of the squares can be used as a 25-point array.*

After lines had been added to act as guides to the eye in counting some unexpected bonuses appeared. The corners of the small squares form 25 points, which can be used as a 25-point graticule if required. Also, the large square can be used as a standard area for cell counts and the like, and various lines can be used for linear intercept counting. It is not suggested that it is as satisfactory for these subsidiary uses as specially designed graticules: nevertheless it can be so used if necessary.

The graticule has been used extensively in the last

six months, and is, I believe, the most satisfactory at present available for most applications of point counting.

My thanks are due to various collaborators, especially Dr K. N. Ahmad and the late Mr W. S. Mack, to C. Jonathan Lennox for checking my mathematics, and to Mr W. D. Hogben of Messrs Graticules for dealing patiently with my exigent requirements.

References

Ahmad, K. N., Dykes, J. R. W., Ferguson-Smith, M. A., Lennox, B., and Mack, W. S. (1971). Leydig cell volume in chromatin-positive Klinefelter's syndrome. *J. clin. Endocr.*, 33, 517-520.

Ahmad, K. N., Lennox, B., and Mack, W. S. (1969). Estimation of the volume of Leydig cells in man. *Lancet*, 2, 461-464.

Dunnill, M. S. (1968). Quantitative methods in histology. In *Recent Advances in Clinical Pathology*, Series V, edited by S. C. Dyke, p. 401-416. Churchill, London.

Dykes, J. R. W. (1969). Histometric assessment of human testicular biopsies. *J. Path.*, 97, 429-440.

Lennox, B., Ahmad, K. N., and Mack, W. S. (1970). A method for determining the relative total length of the tubules in the testis. *J. Path.*, 102, 229-238.

Quenouille, M. H. (1959). *Rapid Statistical Calculations*, p. 6. Griffin, London.