ACTA STEREOL 1983; 2: 337-341
STEREOLOGY AND MORPHOMETRY IN PATHOLOGY
Proc 2nd Symp Morphometry in Morphol Diagnosis
Collan Y. et al., eds.
Kuopio University Press, Kuopio 1983

Reproducibility of stereologic analysis of histologic sections

Marcelino G. Reyes, Thomas Tosch, Valerie Childrey, Clodualdo Penaranda and Collette K. Noland

Department of Pathology, Mount Sinai Hospital, Chicago, Illinois USA 60608

Abstract

In this study, inter-observer and intra-observer errors were compared in 17 stereologic estimates from the counts of five observers on five histologic components of fresh frozen sections of seven muscle biopsies, four histologic components of five carcinomas of the breast and two histologic components of three senile myocardium. Inter-observer error, defined as the difference of the estimates of five observers taken in pairs, had a range of 4 to 56 and 25th, 50th and 75th percentiles of 7, 21 and 28 percent. In contrast, intra-observer error, defined as the difference of two estimates of four observers, had a range of 2 to 12 and 25th, 50th and 75th percentiles of 4, 5 and 8 percent. Our findings indicate that experimental error caused by wrong counts in stereologic analysis of histologic sections can be reduced by using a single observer.

Introduction

Stereology is a statistico-geometric system that is used to predict the quantitative relationships of the three dimensions of structures from their sections that are in two dimensions or less. It has been useful in petrology, metallography (Underwood, 1970) and more recently in biology, particularly in electron microscopy, where cellular organelles have been quantitatively estimated with simple point and line intersection counting technics of stereology (Weibel and Bolender, 1973).

Because it relies on the statistical approach, the major cause of error in stereology is improper sampling. Other errors result from changes in fixation, embedding and cutting of biologic samples. A third cause which has received little attention in histologic sections is wrong counts.

In this study, we compared the difference of stereologic estimates from counts of several observers with the estimates from two counts of each observer to find out the reproducibility of stereologic analysis of histologic sections. Our findings show that reproducibility of stereologic estimates is higher from counts of a single observer than from counts of several observers.

Materials and methods

A. Histologic sections

Seven muscle biopsies, five carcinomas of the breast and three hearts with senile amyloidosis of the left ventricle comprised the study. Biopsies of unfixed muscle were frozen in isopentane at -160°C, cut in serial or step sections at 8 μm and stained with modified trichrome and myosin adenosine triphosphatase (ATPase) at pH 9.4. Carcinomas of the breast were excised at surgery, fixed in formalin, embedded in paraffin, cut in serial or step sections at 5 μm and stained with Verhoef-Van Gieson for elastic tissue and collagen. The hearts were obtained at autopsy, fixed in formalin, embedded in paraffin, cut at 5 μm and stained with crystal violet.

B. Counting technics and stereologic analysis

The muscle biopsies were counted on the microscope using a grid of 25 or 100 squares mounted in the eyepiece. Carcinomas of the breast and senile hearts were counted on a grid of squares painted on a white board where the glass slides were projected by a 35 mm slide projector.

The grids were used for point, line intersection or simple counts. Point counts consisted of the number of points (upper right corner of the square) hitting the histologic components. Line intersection counts were the number of intersections between the boundaries of the profile of the histologic component and the horizontal lines (tops of squares) of the grid. Simple counts were counts of the number of profiles of the histologic components within the frame of the grid or in the section. A mechanical counter was used to record the counts. They were then tabulated and the estimates obtained as follows:

1) *Volume density*: $V/V = P/P$ where P/P is the point fraction or the points hitting the component divided by the total points counted. If the grid size is adjusted by changing objectives or the distance between the slide projector and the board so that the average size of the histologic component is smaller than the squares of the grid, the points hitting that component will be independent and can be considered to be binomially distributed. In this distribution, standard error $SE = \sqrt{[\hat{p}x(1-\hat{p})/n]}$ where $\hat{p}$ is the point fraction P/P and n is the total number of points.

2) *Surface density*: $S/V = 2I/L$ where I is the number of intersections between profile boundaries of the histologic component and the horizontal lines of the grid. The distribution of I in the adjusted test system described was considered as Poisson. In this distribution, standard error $SE = \sqrt{2\,I}/L$.

3) *Numerical density on area*: N/A = number of histologic component divided by area of section or of a histologic component. The latter was the point count multiplied by the square of the length of the side of the squares in the grid. A reciprocal estimate, area of section or of histologic component per component, was also obtained.

4) *Numerical ratios* of histologic components within the frame of the grid or in the section.

C. Observers

Five observers participated in the study: 1)MGR, pathologist who designed the study and set up the counting technics; 2) TT, biostatistician who helped in the design and analysis of the study but had no training or experience in microscopy; 3) VC, research assistant who had practical experience in counting and computing estimates; 4) CP, research technician who counted regularly but had no theoretical knowledge of the estimates; and 5)CKN, histotechnologist, who prepared some of the histologic sections, was experienced in microscopy but not in counting and had no theoretical knowledge of the estimates.

Instructions to the participants were kept at a minimum. On the first day of the experiment, MGR gave a brief outline of the design and purpose of the study. He then made sure that the participants recognized the histologic components and how they were to be counted. Stations were then set up where each of the five participants counted a slide and recorded his count. On the second day of the experiment, the stations were again set up and all participants except TT counted the slide from each station twice. No attempt was made to analyze or to agree on the results of the counts. At the end of the experiments, VC computed the estimates.

D. Statistical analysis

TT analyzed the results using an Apple II Plus computer. Error was defined as one half of the absolute difference between two estimates divided by their average. Inter-observer error was the mean of the errors of the paired estimates of all five observers while intra-observer error was the mean of the errors of two estimates from four observers. Inter-observer error was so defined to be consistent with intra-observer error. Errors were then multiplied by 100 and expressed as percent errors.

The grand mean of the absolute value of each estimate was defined as the average of the mean estimates of each slide. In the inter-observer experiment, the mean of each slide was the average of the estimates of all five observers. In the intra-observer experiment where each slide was counted twice, the mean of the estimate of each slide was the average of two estimates of four observers.

Results

The errors and absolute values of seventeen stereologic estimates from inter-observer and intra-observer experiments are listed in the Table.

Table 1

Errors and values of estimates in inter-observer and intra-observer experiments

Estimates	Error (%) ± se		Values ± se	
	Inter-observer	Intra-observer	Inter-observer	Intra-observer
Muscle biopsies (n = 7)				
Volume density ($mm^3/100\ mm^3$)				
Endomysial collagen	28.4 ± 2.4	6.7 ± 1.1	7.9 ± 0.7	7.4 ± 1.4
Perimysial collagen	16.3 ± 2.2	4.7 ± 0.8	10.1 ± 1.4	7.9 ± 1.3
Capillaries	54.4 ± 10.1	11.5 ± 6.6	1.3 ± 0.2	2.0 ± 0.3
Myofiber nuclei	56.2 ± 8.8	7.8 ± 2.2	2.0 ± 0.4	3.3 ± 0.7
Type 1 fibers	4.4 ± 1.7	1.9 ± 0.3	46.5 ± 3.2	50.9 ± 3.5
Type 2 fibers	5.1 ± 1.7	2.2 ± 0.7	36.6 ± 3.0	36.0 ± 3.7
Interstitium	14.4 ± 1.6	8.7 ± 1.8	16.7 ± 2.3	13.0 ± 1.7
Perimysial collagen	18.0 ± 7.3	5.3 ± 1.7	16.5 ± 3.5	13.6 ± 3.0
Surface density ($mm^2/100mm^3$)				
Capillaries	26.2 ± 5.7	5.1 ± 1.6	780 ± 100	1060 ± 100
Numerical density on area (n/mm^2)				
Capillaries	21.7 ± 2.4	4.2 ± 1.1	592 ± 23.0	647 ± 137
Area / component (um^2/component)				
Myofiber / capillary	21.7 ± 2.4	4.2 ± 1.1	2158 ± 426	2179 ± 494
Numerical ratio (n/100)				
Capillary / myofiber	25.6 ± 1.8	4.9 ± 0.9	18.6 ± 3.3	21.9 ± 4.5
Myofiber nuclei	30.2 ± 2.6	5.8 ± 0.8	30.5 ± 4.4	35.4 ± 4.6
Carcinoma of the breast (n = 5)				
Volume density ($mm^3/100\ mm^3$)				
Elastic tissue	20.6 ± 5.8	9.3 ± 2.3	3.7 ± 2.2	2.2 ± 0.9
Stromal collagen	5.1 ± 1.1	4.8 ± 0.9	56.9 ± 2.5	46.7 ± 4.9
Neoplasm	6.6 ± 1.5	4.2 ± 1.7	39.4 ± 4.1	51.0 ± 4.5
Senile myocardium (n = 3)				
Volume density ($mm^3/100\ mm^3$)				
Amyloid	20.7 ± 3.9	3.5 ± 1.7	6.8 ± 4.0	2.1 ± 1.4

Table shows that the inter-observer errors were larger than the intra-observer errors, although the difference was not significant in the volume densities of type 1 and type 2 fibers in the muscle and of stromal collagen and neoplasm in breast cancer. The lack of significance in these estimates, however, could have been the result of too few observations. In contrast, the absolute values of the estimates expressed as their grand means did not differ in the two experiments.

Ranking of the severity of the errors of the estimates tended to be the same in the inter-observer and intra-observer experiments, i.e., estimates with small, intermediate and large errors tended to be the same in the two experiments. This observation was corroborated by the positive correlation between the inter-observer and intra-observer errors ($r = +.655$, $p<0.01$). Likewise ranking of the absolute values of the estimates was positively correlated in the two experiments ($r = +.806$, $p<0.001$).

Twelve of the estimates were volume densities. Errors among the volume densities tended to be largest in those with the smallest absolute values, e.g., capillaries and myofiber nuclei in muscle and elastic tissue in breast cancer, and smallest in those with the largest absolute values, e.g., type 2 fibers in muscle and stromal collagen in breast cancer. These errors were negatively correlated with the absolute values in the inter-observer ($r = -.751$, $p<0.01$) and intra-observer ($r = -.649$, $p<0.05$) experiments.

We also observed that when we compared each observer's estimate with the mean of the estimates of all observers, each observer tended to operate on one side of the mean or the other. This is shown in Figure 1 for the numerical density on areas of capillaries. Also, the absolute error of an estimate increased with the mean of the estimate as shown on Figure 2 where the lines joining the lowest and highest values diverged as the mean of the slides increased.

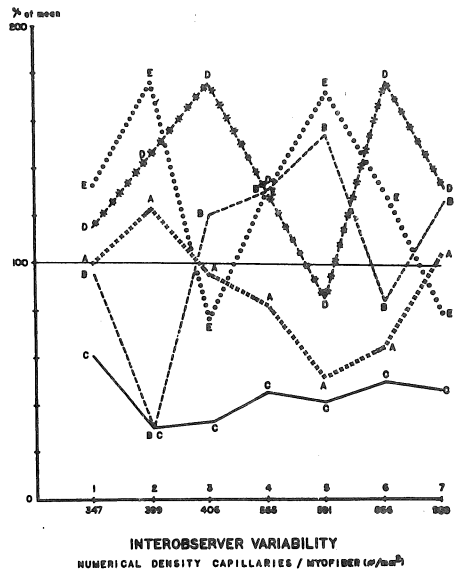

Figure 1. Numerical Density of capillaries on transverse myofiber area. Mean of estimates of five observers for slides 1 to 7 are shown in the abscissa while the ordinate shows estimates of each observer expressed as a percent of the mean. Each observer tended to operate on one side of the mean or the other.

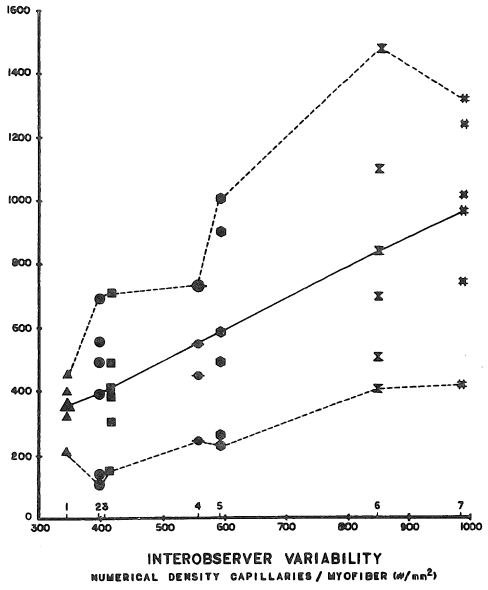

Figure 2. Numerical Density of capillaries on transverse myofiber area. Abscissa shows the mean of the estimates of five observers for slides 1 to 7 while the ordinate shows the absolute value of the estimates of five observers. Solid line shows the means and broken lines the ranges of the estimates of each slide. Absolute errors of each slide increased with the mean of the estimates.

Discussion

At the start of the experiments, we had anticipated that the size of the errors would depend on the quality of the preparations. We expected, for instance, that there will be larger errors in the estimates from counts on the projected images of whole sections of breast carcinomas and of myocardium containing crystal violet deposits because the resolution of these preparations were not as good as those in the microscope. To our surprise, the errors in the volume densities of stromal collagen in breast cancer and of type 2 fibers in muscle which were counted in the microscope were almost identical. Also, we anticipated that the errors on counts on muscle done with the oil immersion objective would be greater than with the low power objective because every observer, inspite of admonitions to the contrary, not infrequently found himself refocusing the specimen when using the oil immersion objective. This led to slight changes in the relationship of the test grid and the section but apparently had little effect on the errors and absolute values of the estimates as shown by the volume density of perimysial collagen which were estimated from oil immersion and low power objectives. In the latter, refocusing was not needed.

Our experience in counting led us to anticipate larger errors in the volume density of rare histologic components. This was shown by the larger errors in the volume density of capillaries and myofiber nuclei compared to endomysial and perimysial collagen of muscle and in elastic tissue compared to stromal collagen and neoplasm of breast cancer. Errors in volume density of rare components, however, can be minimized by adjusting the grid size and/or by increasing the total points counted (Weibel and Bolender, 1973).

Since identification of histologic components was not a problem and since the limitations imposed by resolution, refocusing or clarity of the preparations as well as those inherent in counts of rare histologic components were common to all the observers, we are unable to explain why the inter-observer errors were larger than the intra-observer errors. In a study with experienced histopathologists, Fong *et al* (1979) also found that subjective estimates as well as estimates from point counts of bone marrow aspirates and biopsies were more consistent with an observer than among several observers. It would appear, therefore, that inter-observer errors are larger than intra-observer errors regardless of the expertise of the observers.

We conclude that errors in stereologic analysis of histologic sections caused by wrong counts can be minimized by using a single observer.

References

Fong TP, Okafor LA, Schmitz TH, Thomas W. Westerman MP. An evaluation of cellularity in various types of bone marrow specimens. *Am J Clin Path* 1979;72:812-816.

Underwood EE. *Quantitative stereology.* Reading, Mass., Addison-Wesley Publishing Co., 1970.

Weibel ER, Bolender RP. Stereological technics for electron microscopic morphometry. In: Hayat MA, ed. *Principles and techniques of electron microscopy.* New York, Van Nostrand-Rheinhold Publishing Co. 1973;3:237-296.

Presented in part at the Annual Meeting of the International Academy of Pathology in Boston, Mass., March 2, 1982. This study was supported by grants from Mount Sinai Research Foundation, the Educational Research Foundation of the Association of Pilipino Practicing Physicians of America, the Myra Kurland Heart Fund and the Edward Weinstein Foundation.

We would like to acknowledge Joseph H. Hallissey for his typesetting work on this study.

Reprint requests to MGR, Department of Pathology, Mount Sinai Hospital, Chicago, Illinois USA 60608