ACTA STEREOL 1984; 3/1: 33-38
ORIGINAL PAPER

STEREOLOGIC AND MORPHOMETRIC MEASUREMENTS IN DIAGNOSTIC HISTOPATHOLOGY: STATISTICAL CONSIDERATIONS

Kalevi Selkäinaho and Yrjö Collan

Departments of Applied Mathematics and Pathology, University of Kuopio, P.O.B. 6, SF 70211 Kuopio 21, Finland

ABSTRACT

In terms of variation there are three basic types of studies performed in histopathology (prospective, retrospective and diagnostic studies). There are two ways of applying stereologic or morphometric measurements in such studies: group morphometry/stereology and diagnostic morphometry/stereology. In diagnostic histopathology the sources of variation are more numerous and the range of variation is wider than in prospective or retrospective studies. Also biological variation between normal and pathological samples is larger than between normal samples. This suggests that the principles applied in group morphometry cannot always be recommended in diagnostic morphometry. Test systems in diagnostic histopathology can be best characterized by estimating the reproducibility of the measurements. We have applied kappa statistic and intraclass correlation coefficient for that purpose.

INTRODUCTION

Research in stereology has resulted in reasonably good understanding of the correspondence between 2-dimensional sections and the 3-dimensional space which has been sectioned (Underwood 1970, Weibel 1979). Stereology is based on measurements performed on tissue structures. Such measurements are also characteristic of morphometry which is defined as quantitative description of structure (Weibel 1979). The methods created for the purposes of stereology and morphometry are basically mathematical and probabilistic and do not as such

always pay attention to the special needs or allowances of the fields in which they are applied.

In terms of sources and range of variation there are 3 basic types of studies in histopathology: prospective, retrospective and diagnostic studies. Such studies can be made by applying subjective methods or methods of stereology and morphometry. In a prospective study selection of patients, sampling, fixation, staining, observation and measurements can be well controlled. This means that the potential range of variation due to the type of study is limited. In a retrospective study, selection of patients, sampling or fixation can no longer be controlled well but staining, observation and measurements can be performed with great care. Diagnostic work is done daily in different pathology laboratories in all parts of the world. Numerous pathologists study samples, which are the result of variable fixation, staining, and sampling. Patient selection is not as strictly controlled as in prospective and retrospective studies. If measurements are made variation is to be expected in the methods used. In other words, the diagnostic study has a looser control on study parameters than the prospective study and the retrospective study. The studied population of samples in a diagnostic study could be described as follows: Samples studied within a period of time for diagnostic purposes in different pathology laboratories by different observers after variable fixation and staining, with different microscopes, and different methods of observation and measurement. So the diagnostic study is a continuing study which is influenced by most variable factors in the study environment.

We should realise that diagnostic studies are also carried out in most other fields of human activity, where decisionmaking is crucial.

In histopathology we apply stereology or morphometry in two ways (Collan 1984) . Below we refer to morphometry only. Group morphometry (statistical morphometry) deals with a group of samples and diagnostic morphometry deals with one sample. In the latter type of morphometry our aim is to extract as much valuable information as possible from the sample to make it possible for us to decide about the nature of the disease the patient is suffering from. We base our decision on information which we have gathered through group morphometry (statistical morphometry).

So we have three types of studies (prospective, retrospective, and diagnostic studies) and two ways of applying morpho-

metry in these studies (group morphometry, diagnostic morphometry). In prospective and retrospective studies group morphometry is usually applied. In diagnostic studies we apply group morphometry when we describe the characteristics of our study system, but in our daily work our approach is diagnostic, i.e. we apply diagnostic morphometry.

In group morphometry we now seem to understand the problems of sampling (Cruz-Orive and Weibel 1981, Gundersen and Österby 1981). But we should realise that applications for group morphometry need not necessarily be valuable in diagnostic morphometry. Variation ranges make diagnostic morphometry much different from group morphometry.

AN EXAMPLE

Volume fractions of a tissue component were estimated by point counting from four sections. This was done by applying a a point grid 20 times on these samples. The results (points on the tissue component/points on the whole section) are shown in the following.

Number of measurement	Number of section			
	1	2	3	4
1	1/28	2/26	8/25	13/26
2	1/25	3/26	5/25	9/26
3	1/26	2/30	7/23	9/25
4	1/26	4/25	7/26	11/25
5	1/26	3/26	7/26	14/25
6	0/26	1/25	6/25	10/26
7	1/25	2/25	9/25	10/26
8	0/26	2/26	4/25	15/26
9	1/25	2/25	5/26	8/26
10	2/25	2/26	10/26	10/26
11	1/25	0/25	11/26	10/26
12	2/26	1/25	9/25	10/26
13	2/26	3/25	8/26	16/26
14	0/25	1/25	8/25	12/24
15	2/26	2/24	5/26	13/24
16	2/25	6/25	7/25	11/25
17	0/25	1/25	10/25	12/26
18	2/26	1/26	9/27	12/25
19	2/26	2/26	7/25	11/26
20	2/26	3/27	8/25	15/27

When the results are combined we get the following estimates of the volume fractions (+/-SD):

Number of section	1.	0.047 +/- 0.030
	2.	0.084 +/- 0.052
	3.	0.296 +/- 0.073
	4.	0.451 +/- 0.087

Now we are in the position of thinking how morphometry should be applied to diagnostic pathology. Certain principles are in most instances true in this field and should influence our thinking in this connection.

CHARACTERISTICS OF DIAGNOSTIC HISTOPATHOLOGY

1. Changes in tissues under physiological conditions are usually small, changes between normal and pathological, and between pathological entities vast. If we look how this is reflected in our example we see that a single measurement can make a distinction between sample 1 and samples 3 and 4. If we want to make distinction between samples 1 and 2, or 3 and 4 we need more testing. The former test situation corresponds distinction between normal and pathological samples, the latter distinction between normal samples or between pathological samples.

2. The spectrum of pathological changes is graded. Usually in diagnostic situations we have 3-5 grades. Examples are benign, borderline and malignant tumours in the ovary or the five grades for epithelial changes in uterine cervix. Because these grades are parts of a continuous spectrum it may be difficult and sometimes impossible to make a reliable distinction between various grades with a **single** morphometric measurement because there will necessarily be overlap between grades (Collan 1983). This is especially so because there are several sources of unavoidable variation (in the morphometric method, by human factors, because of biological variation). It is obvious that a clear-cut distinction between grades is possible only when several parameters are applied at the same time in discriminant analysis (Bezemer, Baak, deWith 1977).

3. The histological entities are multifaceted. The various histological entities are defined in terms of several parameters and a correspondence with the classifications need not be reached with a single measurement only. This point, however, should be studied further in each individual problem area. This

also favours the application of several parameters at a time or the combination of several parameters into indexes.

4. Considerable variation is unavoidable in the diagnostic situation. There is vast biological variation, - usually much larger than the variation within physiological states. Human factors cause unavoidable variation. Section thickness and level of focus cause variation.

5. There are practical limits - too laborious measurements cannot be applied.

ESTIMATION OF REPRODUCIBILITY

Rather than trying to apply methods of group morphometry for diagnostic purposes we should characterize the test situation by methods determining the reproducibility of the applied methodology. In graded measurements kappa statistic can be used (Kraemer 1980). Reproducibility for continuous variables can be determined with intraclass correlation coefficient (ICC) (Cochran 1968, Selkäinaho 1983). Our results suggest that reproducibility in the range of moderate - almost perfect can be reached even though the variation e.g. by human factors is considerable (Kosma et al. 1983). The vast biological variation within pathologically changed tissues may also allow a bit more liberal attitude towards sample size than is the practice in traditional group morphometry. However, general rules about this cannot be given and each system applied need be evaluated in its diagnostic context.

REFERENCES

Bezemer PD, Baak JPA, deWith C. Discriminant analysis, exemplified with quantitative features of endometrium. Eur J Obstet Gynec Reprod Biol 1977; 7/3: 209-214.

Cochran WG. Errors of measurements in statistics. Technometrics 1969; 10: 637-666.

Collan Y. Morphometry in pathology: Another look at diagnostic histopathology. Pathol Res Pract 1984; to be published.

Collan Y. Reproducibility in diagnostic pathology and morphometry. In: Collan Y, Aalto M-L, Kosma V-M, Naukkarinen A,

Romppanen T, Syrjänen K, eds. Morphometry and stereology in pathology. Kuopio: Kuopio University Press 1983.

Cruz-Orive L-M, Weibel ER. Sampling designs for stereology. J Microsc 1981; 121: 235-257.

Gundersen HJG, Österby R. Optimizing sampling efficiency of stereological studies in biology: or "Do more less well!". J Microsc 1981; 121: 65-73.

Kosma VM, Collan Y, Aalto M-L, Seppä A, Rautiainen M, Selkäinaho K. Reproducibility and variation in morphometric assessment of positive staining for CEA in ovarian tumours. 6th Int Congr Stereol, Gainesville, Florida 1983.

Kraemer H. Extension of the kappa coefficient. Biometrics 1980; 36: 207-216.

Selkäinaho K. Deriving coefficients of internal consistency of measurements: ICC and kappa. Reports on Statistics, University of Jyväskylä 1983; 12: 1-16.

Underwood EE. Quantitative stereology. Reading, Massachusetts: Addison-Wesley Publishing Company, 1970.

Weibel ER. Stereological methods. Vol.1. New York: Academic Press, 1979.