Science on Form: Proceedings of the First International Symposium for Science on Form,
General Editor: S. Ishizaka, Editors: Y. Kato, R. Takaki, and J. Toriwaki, 533-542.

Diagnostic Morphometry: Relevant Background to Decisionmaking in Diagnostic Histopathology

Yrjö Collan

Department of Pathology, University of Kuopio, P.O.B. 6, SF-70211 Kuopio, Finland

Keywords: Diagnostic histopathology, morphometry, stereology, image analysis

Histopathology is a morphological science which has created a disease classification (investigative histopathology) and uses it to identify disease cases (diagnostic histopathology). Identification is based on the decision by the observer (pathologist). Recently morphometric methods have been introduced to help in diagnostic decisionmaking. One of the main lines of investigation is the search for good sampling rules. Also in morphometric histopathology distinction should be made between investigative and diagnostic morphometry. The latter usually involves a greater number of variation sources than investigative morphometry. This means that - for the same level of accuracy and at the same level of probability - more measurements may be necessary in diagnostic morphometry than in investigative morphometry .

INTRODUCTION

Histopathology as a morphological science can be compared with ornithology. In the latter, lots of scientific efforts have been directed to creation of a biologically relevant classification of species. Today this classification is quite covering and includes more than 8900 species (Clements: 1978). It is not very probable that many new species are found to be added to the list, only a few additions can be expected. One of the latest additions was Okinawa rail in Japan (Wild Bird Society of Japan: 1982). Classification of species also includes characterisation of the features which have turned out to be most relevant for identification of individual birds in the field. Here one should make a distinction between classification and identification. The former process results in the Classification Table of Birds. The latter process identifies individual birds seen on the field and does that by trying to find a proper place in the Classification Table of Birds for each bird seen. So classification and identification are two separate intellectual processes, although the term classification in an unnecessarily liberal fashion has been used to cover any or both of the two.

Histopathology is dualistic in a corresponding way. Classification of diseases is based on a variable collection of biologically relevant data on disease: causative organisms, pathogenetic mechanisms, clinical pattern of symptoms and signs,

histological findings, chemical laboratory data, response to treatment with a drug. Also sociological aspects (e.g. in alcoholism), tradition, available methods, and recent research findings affect disease classification as it is used in various hospitals around the world. Disease classification is far less perfect than classification of bird species. The classification has often changed and changes are to be expected in the future. It should also be realised that application of the disease classification varies in line with the variation of the incidence of diseases in various places.

One of the most recent additions to the disease classification is AIDS (acquired immune deficiency syndrome). This disease was not known before 1980. The disease was first identified by the symptom complex it caused. Later the symptom complex, and a specific virus associated with it, have identified the condition.

Microscopic histopathology has been in a key role in creating disease classification. This is why the histopathological (histological) findings associated with most disease conditions are well known. This also applies to diseases which are primarily characterised by other than histological criteria, e.g. AIDS. The morphological classification of diseases is the basis of graduate education in pathology, and is the system applied for the identification of diseases in diagnostic histopathology. The histological disease classification is especially important in the identification of tumors, but also in the identification of other conditions. When there is a problem, solution is often looked for through histopathology - with the idea of getting a clue for further action, if not the diagnosis.

So, we can speak of investigative pathology which creates classifications, and diagnostic pathology which identifies disease conditions.

DIAGNOSTIC DECISION

When a disease condition is identified in the diagnostic context, physicians speak of making a diagnosis. It would be much more appropriate to speak about a diagnostic decision. The latter expression makes it immediately clear that we are not dealing with exact concepts in the traditional scientific sense. Rather, we have a situation in which the decisions made by various physicians can vary. The diagnostic decision between benign and malignant tumors is usually highly reproducible (Linden et al.:1960). The same is not, however, true of decisions on the malignancy grades of tumors (Ringsted et al.: 1978, Collan: 1982, Collan: 1983).

Another point which is important in this connection deals with the accuracy of the decision. Very often the physician is in a situation in which there are no ways of defining the correct decision about the diagnosis, because no absolute reference source is available. In deciding between benign and malignant the absolute reference information may be available after a follow-up of several years. If during that period there is a metastasis from the tumor, the tumor should be considered malignant. The diagnostic decision usually cannot allow a follow-up because

proper treatment should be started as soon as possible. The latter point forces the pathologists to use special methods in addition to traditional histopathology: special stains, immunohistochemistry, cytophotometry for DNA, tissue culture, chromosome analysis, flow cytometry, morphometry and other types of quantitative histopathology. Here we are especially interested on morphometric parameters. In the following I will mostly concentrate on diagnostic morphometry of tumors.

MORPHOMETRIC PARAMETERS IN HISTOPATHOLOGY

The sections studied in histopathology are 2-dimensional. In the light of the studies which so nicely have correlated the 3-dimensional structure of tissue with its function (Bolender 1983, Weibel 1983) we have to ask ourselves whether it would be advisable to try to apply 3-dimensional (stereological) parameters also for diagnostic decisionmaking.

There are certain contexts in histopathology, in which one necessarily should characterise the tissue in 3-dimensional terms. If one aims at volumetric measurements of tissue components, or at parameters per tissue volume, at fresh tissue values, at correlation of biochemical and morphometric data, at the 3-dimensional structure of tissue or its components, or is involved with studies in which volume changes take place, stereological parameters are to be recommended because parameters measured from the 2-dimensional section do not reflect the changes reliably in 3 dimensions. In the diagnostic context, however, the stereological parameters are not always necessary, and in fact, most of the relevant data so far collected on diagnostic histopathology of tumors are on primary measurements on the 2-dimensional sections. Here the main question has been the distinction of two types of lesions, or finding a parameter which is correlated with prognostic aspects of the tumor under study. In the following I will limit my presentation on a few studies in which the morphometric parameters have been well correlated with the aspects of prognosis.

In the histopathology of tumors, morphology of cell nuclei has always had a key position. Measurements most often made involve nuclear area, diameter, perimeter, and form factors. With measurements of these features various types of lesions can be distinguished (Kosma et al.: 1986). But the mean values of these parameters are only one aspect of the data so collected. Also the variation and other features of the distribution are important. These can be analysed by calculating the standard deviation or other moments of the distribution (Bartels: 1979, Longley-Cook: 1971). The predictive power of the standard deviation over the mean value was demonstrated by Gamel et al. (1982) in their study on uveal melanoma. This rule, however, is not necessarily applicable to any tumor. Baak et al. (1982) showed, in carcinoma of the breast, that nuclear perimeter was a significant prognostic predictor but that the standard deviation of the nuclear perimeter was not. Nuclear area and perimeter measured on the section have shown good correlation with prognosis in breast carcinoma and uveal melanoma, but have also been valuable in distinguishing malignant and benign lesions (Bhattacharjee et

al.: 1985), and different types of benign lesions (Kosma et al.: 1986). One type of computerised (Oja & Collan: 1983) morphometry, densitometry, gives basicly the same information as DNA microphotometry, which has been shown to give important prognostic predictors (Atkin & Kay: 1979).

Morphometric parameters on nucleoli have been measured with automatic (Voss et al.: 1983) or semi-automatic methods (Gamel et al.: 1982, Baak: 1985), and their importance as prognostic predictors have been demonstrated in uveal melanoma (Gamel et al.: 1982) and breast carcinoma (Baak et al.: 1985).

The fraction of cells in DNA synthesis in a tumor is an important prognostic predictor (Tubiana et al.: 1984). Correspondingly the fraction of cells in mitosis (mitotic index) has been shown to be a good predictor of prognosis (see e.g. Baak et al.: 1981, and Baak et al.: 1982). Mitotic index measured on the 2-dimensional section is a good estimate of the mitotic index in the 3-dimensional tissue.

Cellularity index is usually measured as the number of cells seen within a unit area of the image. Because breast cancer cells were larger than benign cells, survivors had a higher cellularity index than non-survivors (Baak et al.: 1982). The cellularity index measured on a section is not as such a good estimate of the cellularity index in 3-dimensional tissue. To estimate a cellularity index for the 3-dimensional situation one has to know the shape or the caliper diameter of the nuclei or cells, or the measurements have to be made from two sections with different thickness (Ebbeson & Tang: 1967, Collan et al.: 1983).

Volume fraction of the epithelium, volume fraction of the glands in the tumor, and area of the luminal surface of the glands per volume of tissue, have been shown to be of diagnostic and prognostic significance in endometrial lesions (Baak et al.: 1981). These parameters are derived from measurements on the 2-dimensional section but they describe the 3-dimensional situation. Such 3-dimensional approach may look especially valid in morphometric histopathology. However, most of the data available are on the 2-dimensional situation. This is in line with the tradition in histopathology: histopathologic interpretation has been based on the 2-dimensional microscopic image. Vast amount of information has been collected from such images, and associated with prognostic, therapeutic, diagnostic, and other clinical data. This collection of information has been going on for years and it has also thoroughly influenced disease classifications. Much of the data has been qualitative in nature. Now that we would like to associate quantitative aspects of the image with the already available information pool, it is quite natural that we start from the 2-dimensional image and parameters measured from that image. If the stereological approach gives additional advantages, it should be applied, of course.

What approach is taken depends - in the final evaluation - on medical relevance, which includes many aspects. Of these, two are more important than the others. The method and parameters measured by the method should be adaptable to the medical diagnostic system - to the culture of diagnostic histopathology. If this is not possible, the method cannot reach the level of popularity which is necessary for general acceptance. Second, of

the methods technically applicable, those are chosen which give the best predictive power in respect to presence of disease, to survival, to response to specific treatment, or to other aspects important for the patient. Single predictors can be combined for better correlation. The theory or philosophy behind the technical performance is not important here. So, from the standpoint of medical decisionmaking it is irrelevant whether the method applies parameters related to 3 dimensions, or parameters related to 2 dimensions.

SAMPLING RULES

It is not difficult to understand the obvious need for the application of morphometric methods in diagnostic histopathology (Collan: 1982). But how should they be applied? There are no definite answers, and especially there are no general or generally accepted rules for sampling. Baak & Oort (1983) advocated selective sampling: the pathologist should define the area to be measured. Baak and his group have got excellent results with this approach. A basicly different approach tries to find a representative mean value for the sample. How morphometry is applied should be defined clearly, so that the measurements can be reproduced. In a recent meeting morphometric results on bladder carcinoma were reported by 4 groups. These groups used different sampling rules (Helander et al.: 1985, Mariuzzi et al.: 1985, Nielsen et al.: 1985, Veldhuizen et al.: 1985). Good sampling rules are especially important for the development of morphometric pathology.

A simple example will show the point. Microscopic fields with 40 to 50 nuclei were selected from 3 different lesions. The main idea here was to demonstrate the variability of the results. Nuclear areas were measured and the mean value and standard deviation were recorded. The results were as follows:

Carcinoma of the prostate	114 +- 23,	122 +- 24,	105 +- 22
Adenocarcinoma of the kidney	72 +- 17,	98 +- 17,	71 +- 20
	78 +- 13,	80 +- 28,	70 +- 17
Lingual fibromatosis	102 +- 27,	91 +- 29,	59 +- 27
	38 +- 11		

It is easy to see that the mean values differ in different fields. How to select the most relevant field? This question still waits its answer in most fields of histopathology.

DIAGNOSTIC MORPHOMETRY

The importance of making a difference between investigative and diagnostic morphometry was stressed in recent articles (Selkäinaho & Collan: 1984, Collan et al.: 1986). If this distinction is made, also diagnostic histopathology routine can be approached in scientifically sound fashion. The distinction includes a realistic attitude towards the many variation sources present in diagnostic morphometry. In diagnostic histopathology routine the samples are studied in different laboratories in different parts of the world, by different observers, after different kinds of fixation, sectioning and staining. There are

numerous variation sources. To be able to apply morphometry in the diagnostic situation one should know the system in which one works. This means that we should know the size of the interlaboratory variation, the interobserver variation, the intraobserver variation, the interfield variation, and (in nuclear measurements) the internuclear variation. Estimates of the size of these variation sources have been given. In a study the nuclear areas on section were considered (Collan et al.: 1986). It turned out that the standard errors in a system with interobserver sources of variation were distinctly larger than in a system without interobserver sources of variation:

Type of system	SE	Variance
System with interobserver variation	3.5	12.4
System with intraobserver variation only	2.6	6.8

It is easy to see that the variation in a system with interobserver variation is about double the variation in a system with intraobserver variation only. The situation, however, can be handled by statistical methods. By applying the the confidence limits (at a defined level of probability, here 95%):

$$x \pm 1.96 \frac{SD}{\sqrt{n}}$$

(x = mean of the measurements, SD = standard deviation of the measurements, n = number of nuclei measured) we can estimate the number of nuclei we should measure in each of the two situations. With a mean nuclear area value of 47.7 square micrometers, and confidence limits within +- 10% of the mean, the following numbers of nuclei should be measured (Collan et al.: 1986):

System with interobserver variation	93 nuclei
System with intraobserver variation only	50 nuclei

It is obvious that morphometric classification should be based on research in which the variation sources are small (systems with intraobserver variation only). On the other hand, a realistic view of diagnostic morphometry should include consideration of all variation sources involved.

REFERENCES

Atkin, N.B., Kay, R. (1979): Prognostic significance of modal DNA value and other factors in malignant tumours, based on 465 cases. Br. J. Cancer 40: 210 - 211.

Baak, J.P.A. (1985): The relative prognostic significance of nucleolar morphometry in invasive ductal breast cancer. Histopathology 9: 437 - 444.

Baak, J.P.A., Agrafojo Blanco, A., Kurver, P.H.J., Langley, F.A., Boon, M.E., Lindeman, J., Overdiep, S.H., Nieuwlaat, A., Brekelmans, E. (1981): Quantitation of borderline and malignant mucinous ovarian tumours. Histopathology 5: 353 - 360.

Baak, J.P.A., Oort, J. (1983): Introduction to the applications. A Manual of Morphometry in Diagnostic Pathology. (ed. J.P.A. Baak and J. Oort. Berlin. Springer-Verlag.): 48 - 51.

Baak, J.P.A., Kurver, P.H.J., de Snoo-Nieuwlaat, A.J.E., de Graef, S., Mankkink, B., Boon, M.E. (1982): Prognostic indicators in breast cancer - morphometric methods. Histopathology 6: 327 - 339.

Bartels, P.H.(1979): Numerical evaluation of cytological data. I. Description of profiles. Analyt. Quantitat. Cytol. 1: 20 - 28.

Bhattacharjee, D.K., Harris, M., Faragher, E.B.(1985): Nuclear morphometry of epitheliosis and intraduct carcinoma of the breast. Histopathology 9: 511 - 516.

Bolender, R.P. (1983): Integrating methods: A key role for stereology. Acta Stereol. 2: Suppl.1: 131-138.

Collan, Y. (1982): Reproducibility, the neglected cornerstone of medical diagnostics. Morphometry in Morphological Diagnosis. Ed. Y. Collan, and T. Romppanen. Kuopio. Kuopio University Press.

Collan, Y. (1983): Reproducibility and sources of variation in diagnostic histopathology and in diagnostic morphometry. Acta Stereol. 2: 298-307.

Collan, Y., Mariuzzi, G.M., Torkkeli, T., Montironi, R., Marinelli, F., Collina, G., Kosma, V.-M., Pesonen, E., Kosunen, O., Jantunen, E. (1986): Sampling in diagnostic morphometry. The influence of variation sources. Path. Res. Pract. (in press).

Collan, Y., Oja, E., Whimster, W.F. (1983): Mathematical background to stereology and morphometry for diagnostic pathologists. Acta Stereol. 2: 214 - 238.

Clements, J.F.(1978): Birds of the World: a Check List. The Two Continents Publishing Group. New York.

Ebbeson, S.O.E., Tang, D.B. (1967): A comparison of sampling procedures in a structured cell population. Stereology. (ed. H. Elias. Berlin. Springer.): 131 - 132.

Gamel, J.W., McLean, I.W., Greenberg, R.A., Zimmerman, L.E., Lichtenstein, S.J.: Computerized histologic assessment of malignant potential: A method for determining the prognosis of uveal melanomas. Human Pathol. 13: 893 - 897.

Helander, K., Johansson, S., Nielsson, S.(1985): Urinary bladder carcinoma grade II: Cytometry and morphometry. Path. Res. Pract. 180: 275.

Kosma, V.-M., Collan. Y., Naukkarinen, A., Aalto, M.-L., Männistö, P. (1986): Histopathological and morphometrical analysis of skin changes in NMRI mice induced by dithranol

(anthranil) and its acyl analogs. Arch. Toxicol. : in press.

Linden, G., Cline, J.W., Wood, D.A., Guiss, L.W., Breslow, L. (1960): Validity of pathological diagnosis of breast cancer. JAMA 173: 143-147.

Longley-Cook, L.H.(1971): Statistical Problems and How to Solve Them. Barnes & Nobel Books, New York.

Mariuzzi, G.M., Montironi, R., Scarpelli, M., Pisani, E., Ansuini, G.(1985): Classification of non-invasive urothelial papillary lesions by the malignancy index. Path. Res. Pract. 180: 294.

Nielsen, K., Colstrup, H., Nilsson, T.(1985): Morphometric investigations on urinary bladder carcinomas, correlated to histopathological grading and to prognosis. Path. Res. Pract. 180:302.

Oja, E., Collan, Y.(1983): Basic principles of image analysis by a computer. Acta Stereol. 2: 250 - 261.

Ringsted, J., Amtrup, F., Asklund, C., Baunsgaard, P., Christensen, H.E., Hansen, L., Jakobsen, C., Jensen, N.K., Moesner, J., Rasmussen, J., Reintoft, I., Rolschau, J., Starklint, H., Thommesen, N., Vrang, J. (1978): Reliability of histo-pathological diagnosis of squamous epithelial changes of the uterine cervix. Acta Pathol. Microbiol. Scand. A 86: 373-278.

Selkäinaho, K., Collan, Y.(1984): Stereologic and morphometric measurements in diagnostic histopathology: statistical considerations. Acta Stereol. 3: 33-38.

Tubiana, M., Pejovic, M.H., Chavandra, N., Contesso, G., Malaise, E.P. (1984)): The long-term significance of the thymidine labelling index in breast cancer. Int. J. Cancer 33: 442 - 445.

Veldhuizen, R.W., Boon, M.E., Ooms,, E.C.M. (1985): Histomorphometry of transitional cell tumours. Path. Res. Pract. 180: 322.

Voss, K., Hufnagl, P., Martin, H., Roth, K. (1983): Karyometric data by image analysis and their use in pathology. Acta Stereol 2: 312 - 318, 1983

Weibel, E.R. (1983): Stereology in biological morphometry: Evaluation of scope and sophistication over two decades. Acta Stereol. 2: Suppl.1: 211-218.

Wild Bird Society of Japan, Massey, J.A., Matsui, S., Suzuki, T., Swift, E., Hibi, A., Ichida, N., Tsukamoto, Y., Sonobe, K., Takana, S., Monna, N., Robinson, J.W. (1985): A Field Guide to the Birds of Japan. Kodansha International. Tokyo.

12-6
Q: 1) I would like to make a general statement: "Measuing 3-D parameters will increase your power to discriminate when compared to 2-D parameters". I would appreciate your comments.

2) You mention the number of nuclei that need to be measured within each patient under various conditons without mentioning the inter-patient (i.e. biological variance). Surely it is only meaningful to calculate intra-individual sample sizes when set in the overall contact of the biological (i.e. inter animal) variance? (V. Howard)

A: 1) All depends on what you want to discriminate. I would like to consider discrimination between diseases first. As I mentioned, disease classification is based on many kinds of features, including histopathology. However, histopathology as such is not able to distinguish all disease entities, although we know what kind of changes can be expected in each condition. There are diseases with identical histopathology, but with different etiology. For instance, we cannot usually tell the cause of fatty liver just by looking at the microscope section. Other information is also necessary. In morphometric histopathology situations with identical results do not allow distinction between two disease conditions even though their etiology, and pathogenesis may be so much different that are classified as two separate entities. Under such circumstances the dimension of the space in which the measurements are made is irrelevant.

When we deal with discrimination between different disease entities we have to realize that the changes on which we base the diagnostic decision may be histological artifacts - e.g. the ground-grass nucleiin papillary carcinoma of the thyroid gland, or lacunar cells in a type of Hodgkin's disease. There may be features of corresponding nature also in the morphometric approach, features that do not seem to follow the path of logical thinking, but simply turn out to be most helpful. There can, of course, be measurements either in the 2-dimensional or 3-dimensional space.

About discrimination between diseases I would like to repeat what I said in my lecture. The final decision about the validity of a measurement in histopathology is done on medical grounds. This has not always been understood. The approach that we have in anatomy, in which the 3-dimensional structure of tissues is important, is not always repeated in histopathology. The main idea is not to get a correct 3-dimensional image of the tissues, but information that can - in a reproducible fashion - be related with clinically relevant parameters. Just think yourself as a patient and you understand immediately where the stress should be.

Finally, I would like to come to the point of your question. Now, I speak about the structure of tissue only. Not on discrimination between diseases, but on discrimination of two tissue samples from each other. The stereological image, which describes the true structure of tissues in 3 dimensions is, of course, the goal we should aim at when we interprete findings on 2-dimensional sections. I would like to make it clear that I have nothing against this principle. Your statement, however, is true only in the ideal situation in which we really have a truthful 3-dimensional image available. The practical situation is different. In diagnostic histopathology serial sectioning is

out of question in most cases. We have to make measurements on section on a 2-dimensional space, and then derive the parameters of the 3-dimensional space. Often after approximations. When we measure the features in the 2-dimensional space to derive the parameters of the 3-dimensional space, we usually sample the image (line grids, point grids etc.). All of this means that there are additional sources of variation in the derived parameters. Such variation may make discrimination problematic, or cause the need for extra measurements. If other conditions are equal, of parameters characterising the samples those should be chosen for diagnostic purposes which give a discrimination with the least amount of effort. By their nature (e.g. volume being the third power of length) parameters of the 3-dimensional space may give certain advantages in discrimination. However, they start to lose this advantage if the estimation cannot be made as accurately as the estimation (=measurement) of the parameters in 2-dimensional space.

After this it is quite easy to comment your statement. In the ideal situation in which the 3-dimensional structure of tissue is known accurately (as in a model of the exact sciences) your statement is certainly true in many if not in most situations. In practice, however, the influence of the variation sources involved cannot be predicted in a general way. The truth value of the statement can only be estimated through testing the discrimination capacity of various parameters in each study situation.

2) I was speaking about the accuracy of measurement. What you have taken up deals with something else. The differences of our approaches seem to be related to the differences between group (=investigative) and diagnostic morphometry. I compared the variation in two systems, which measured the same sample. There were numerous observers. One observer also measured the sample many times. What you speak about concerns numerous samples, but you are not interested in the number of observers. For each diagnostic situation the accuracy of measurements should be defined and this is related to the classification used for the diagnostic purpose. There may be some differences between samples (transitional cell carcinoma, Grade II) which were studied. For any other kind of investigation the numbers of measurements should be estimated separately, in accordance with the needs of accuracy, and probability of occurence of the defined accuracy, in each study. The numbers of measurements necessary are also dependent on the type of the sampling rule applied.

Q: Do you mean 3-dimensional parameters derived from 2-dimensional data has not so much predictive power comparing to 2-dimensioal parameters? (T. Togawa)

A: No, such general rule cannot be given to data, however, we do not have data on the prognostically important 3-dimensional parameters. The only relevant research has been done on 2-dimensional parameters. Of these, of course, 3-dimensional parameters can be calculated. There is no doubt, however, about the fact that in the picture many 3-dimensional parameters will turn out, to be important as good prognostic predictors.