

OBSERVER VARIATION AND REPRODUCIBILITY OF GRADING: ANALYSIS OF
THE POSTCAPILLARY VENULES IN HUMAN AXILLARY LYMPH NODES USING
SUBJECTIVE AND MORPHOMETRIC METHODS

Veli-Matti Kosma, Yrjö Collan, Kari Syrjänen,
Marja-Leena Aalto, Anders Seppä,
and Kalevi Selkänaho*

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

ABSTRACT

Five pathologists analysed 100 postcapillary venules (PCVs) in 20 axillary lymph nodes by subjective and morphometric means. The coefficient of variation (CV) of the subjective mean grade estimates and indexes varied from 0% to 39% and of the morphometric mean height estimates and indexes from 2% to 44%. The interobserver reproducibility was estimated by the kappa statistic and the intraclass correlation coefficient (ICC). On subjective grading kappa was 0.50 (0.43-0.56) and on morphometry ICC was 0.74 (0.70-0.78). These figures correspond to moderate degree of reproducibility, and show that the subjective and morphometric methods are comparable in their performance in this system. Experience in diagnostic histopathology did not appear to have any influence on reproducibility as shown by an exclusion experiment. This suggests that experience amplifies other than basic perceptive or observation skills of the pathologist.

INTRODUCTION

Postcapillary venules with high cuboidal endothelium are found in the paracortex of the lymph nodes in man and other species (Cottier et al., 1973). In recent years attention has been focused on the structural changes of the endothelium of PCVs depending upon the activity of local immune response

(Kittas and Henry 1981, Syrjänen 1982). Studies in man have shown that the higher the endothelium of the PCVs in axillary lymph nodes the better is the prognosis with respect to metastasis in breast carcinoma patients (Syrjänen 1982). So, it is valuable to grade postcapillary venules. Most grading systems apply subjective grading. From diagnostic point of view it is important to know to what extent subjective results parallel morphometric measurements.

In this study we compared the subjective and morphometric measurements on lymph node postcapillary venules. Also we evaluated the reproducibility of grading by applying the kappa coefficient and the intraclass correlation coefficient (ICC).

MATERIALS AND METHODS

The study was based on a series of 20 axillary lymph nodes from 17 patients dissected in association with mastectomy in Kuopio University Central Hospital, Kuopio, Finland during the years 1980-81. The lymph nodes were fixed in 10% buffered formalin and embedded in paraffin. Five micrometer thick histological sections were cut and stained with hematoxylin and eosin. Five pathologists estimated the subjective grade of the same five selected postcapillary venules in each of these 20 lymph nodes at a magnification of 63 (objective lens) x 10 (eye piece) and thereafter measured the thickness of the endothelium with an eye piece graticule (El, Graticules Ltd., England).

The subjective grading was based on the following criteria (Syrjänen 1982):

- Grade 1: Postcapillary venules had flattened endothelium, and showed no or few intraluminal lymphocytes.
- Grade 3: The endothelium was high and cuboidal and there were numerous intraluminal lymphocytes.
- Grade 2: Intermediate between grades 1 and 3.

On morphometry the dimensions measured were: D_{pcv} , the diameter of the postcapillary venule, which was measured at right angles to the longer semiaxis at the level of the smaller semiaxis. D_{lu} , the corresponding diameter of the lumen was also recorded. The height of the endothelium, H_{end} , was then calculated from the equation:

$$H_{end} = \frac{D_{pcv} - D_{lu}}{2} \quad (\text{Fig. 1})$$

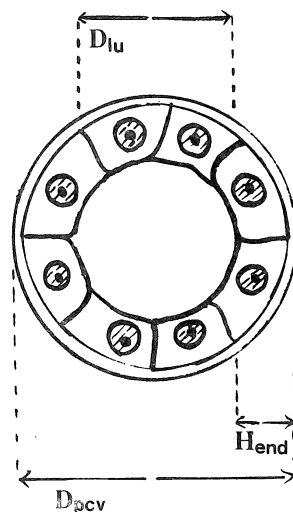


Fig. 1

After measurements the following parameters were calculated: (1) The mean grade estimate, the mean of five grading estimates of the same vessel, grading performed by 5 observers. (2) The grade index, the mean of grades of five vessels of the same lymph node as graded by a single observer. (3) The mean grade index, the mean of five grade indexes, grading performed by 5 observers. (4) The mean height estimate, the mean of five Hend values of the same vessel, grading performed by 5 observers. (5) The height index, the mean of Hend values of five vessels of the same lymph node graded by a single observer. (6) The mean height index, the mean of five height indexes of the same node, measurements performed by five observers. For statistical analysis the standard deviations and the coefficients of variation (CV) of the above parameters were calculated. Pearson's r correlation coefficient was calculated for the estimates and indexes. Also the interobserver reproducibility was analysed by applying the kappa coefficient (Landis and Koch 1977, Kraemer 1980, Selkänaho 1983) and the intraclass correlation coefficient (ICC) (Cochran 1968, Selkänaho 1983).

RESULTS

In Table 1 there is a random sample of the coefficients of variation (CV) of the results of subjective grading and morphometric measurements of the individual postcapillary venules done by 5 pathologists. The CV varies from 0% to 39% in subjective grading, and from 3% to 44% in morphometry. The corresponding means and standard deviations of CV are 0.17 \pm 0.13 and 0.17 \pm 0.09, respectively.

Table 2 shows the coefficients of variation of the mean grade and height indexes of subjective grading and morphometric measurements from the individual lymph nodes. In subjective grading the results vary from 0% to 20% and in morphometry from 2% to 21%. The corresponding means and standard deviations are 0.11 \pm 0.05 and 0.09 \pm 0.05, respectively.

Table 3 gives the interobserver reproducibility of results, estimated with the kappa coefficient and the intraclass correlation coefficient (ICC). Both in subjective and morphometric measurements the results (0.50 and 0.74) correspond to moderate degree of reproducibility. Also in measurements by 4 observers the results were similar. However, there was variation in the range of measurements of different observers in respect to the mean values, but the grade of reproducibility remained the same.

Pearson's correlation coefficients among 5 observers in case of individual venules vary from 0.62 to 0.89 and in case of individual lymph nodes from 0.65 to 0.93. 90% confidence limits among observers vary from 0.50 to 0.92 and from 0.36 to 0.97, respectively.

Table 1. The variation in subjective grading and morphometric measurements of individual postcapillary venules. The venules were selected at random from the 100 postcapillary venules estimated in this study. Estimates were made by 5 pathologists. The results are expressed as coefficients of variation (standard deviation of the estimates divided by the mean). The means shown were calculated from 100 vessels.

Venule number	CV of individual mean grade estimates (subjective)	CV of individual mean height estimates (morphometric)
2	0	0.09
3	0.34	0.29
4	0.39	0.44
10	0	0.12
11	0	0.16
17	0.20	0.04
24	0.25	0.25
31	0.20	0.13
45	0.25	0.17
47	0.20	0.11
57	0	0.15
59	0.25	0.11
60	0	0.16
61	0.37	0.28
63	0.37	0.23
75	0.20	0.10
76	0	0.18
79	0.23	0.11
81	0.20	0.10
83	0	0.23
88	0.39	0.20
91	0	0.33
95	0	0.30
99	0.23	0.07
100	0.39	0.37
	Mean 0.1675 +- 0.1340	Mean 0.1654 +- 0.0887

DISCUSSION

The decisionmaking in diagnostic histopathology is based on strict morphologic criteria but there is also a certain amount of intuition in the process. So it is reasonable to bring to attention the importance of inconsistencies in diagnosis between the pathologists as shown in several studies (Cocker et al., 1968, Lambourne and Lederer 1973, Sissons 1975, Reyes et al., 1983). Statistical parameters have been

Table 2. The variation in subjective grading and morphometric measurements of postcapillary venules in 20 lymph nodes. From each lymph node 5 venules were selected for subjective and morphometric measurements.

Number of lymph node	CV of the mean grade indexes (subjective)	CV of the mean height indexes (morphometric)
1	0.12	0.17
2	0.11	0.10
3	0.07	0.03
4	0.12	0.12
5	0.14	0.05
6	0.08	0.02
7	0.18	0.06
8	0.12	0.09
9	0.13	0.17
10	0.16	0.06
11	0.06	0.21
12	0.04	0.12
13	0.20	0.17
14	0.12	0.06
15	0.05	0.04
16	0.11	0.08
17	0.06	0.05
18	0.09	0.06
19	0	0.10
20	0.15	0.06

Mean: 0.1055 +- 0.0496 Mean: 0.0910 +- 0.0530

Table 3. Interobserver reproducibility as estimated with the kappa coefficient and the intraclass correlation coefficient (ICC). Kappa and ICC were calculated for grade estimates and height estimates of individual postcapillary venules, measured by 5 observers. The same parameters were also calculated for 4 observers, each pathologist being excluded in turn. 90% confidence limits are indicated in brackets.

	Individual venules		Excluded pathologist
	Subjective grading Kappa	Morphometric measurements ICC	
5 measurements	0.50*** (0.43-0.56)	0.74*** (0.70-0.78)	
4 measurements	0.59*** (0.43-0.74)	0.76*** (0.72-0.80)	1
4 measurements	0.44*** (0.37-0.51)	0.71*** (0.67-0.76)	2
4 measurements	0.49*** (0.42-0.56)	0.77*** (0.73-0.81)	3
4 measurements	0.56*** (0.41-0.72)	0.74*** (0.69-0.78)	4
4 measurements	0.56*** (0.49-0.63)	0.73*** (0.68-0.77)	5

*** moderate degree of reproducibility

developed for the analysis of the reproducibility of histological measurements by different observers (Kraemer 1980, Selkäinaho 1983, Silcocks 1983). Morphometry should be able to increase accuracy of interpretation and in theory should be more reliable than subjective grading systems in the analysis of histological parameters. Morphometry should also decrease the variation in measurements by different pathologists.

This need not always be true, however. In our study the variation (in terms of coefficient of variation) was similar in both methods (Tables 1 and 2). Also interobserver reproducibility between 5 pathologists was almost identical in morphometry and in subjective grading. In both situations the results corresponded to moderate degree of reproducibility. Also kappa and ICC values corresponded to same degree of reproducibility in measurements by 4 pathologists. The latter approach allowed the analysis of the influence of experience in these measurements. In subjective grading kappa was identical (0.56) when the most senior or the most junior pathologist was excluded. Also in morphometry the values were quite similar: 0.73 (senior pathologist excluded) and 0.74 (junior pathologist excluded). Those values corresponded to moderate degree of reproducibility. There was also a large variability in the values of Pearson's r correlation coefficients among 5 observers in case of PCVs and lymph nodes (Kosma et al., 1984).

Our results suggest that it is not always necessary to apply morphometric methods if simple grading systems are alternatives which work well in practice. The test system was quite simple and in more complicated subjective grading systems reproducibility might be less good. In this simple system morphometric measurements do not seem to increase reproducibility markedly. This is partly due to the fact that measurements based on eye piece graticules still have a subjective component which causes variation. An alternative method could be based on photography, which could give better reproducibility. However, sources of variation are inherent in the latter method also because selection of focusing levels is variable - different observers could select different focusing levels. One should realise that the reproducibility in subjective grading may be superior even though the results can be biased. The bias was not estimated in this study. It is probable that bias causes variation in subjective grading. Bias is not as prominent in morphometry, and the means of the results in morphometric measurements should approach the real values when the number of measurements increases.

REFERENCES

- Cochran WG. Errors of measurements in statistics. *Technometrics* 1968; 10: 637-666.

- Cocker J, Fox H, Langley FA. Consistency in the histological diagnosis of epithelial abnormalities of the cervix uteri. *J Clin Path* 1968; 21: 67-70.
- Cottier H, Turk J, Sobin L. A proposal for a standardized system of reporting human lymph node morphology in relation to immunological function. *J Clin Path* 1973; 26: 317-331.
- Kittas C, Henry L. BCG-induced changes in the post capillary venules of the guinea pig lymph node. *Lymphology* 1981; 14: 24-28.
- Kosma V-M, Selkänaho K, Collan Y, Syrjänen K, Aalto M-L, Seppä A. Grading of postcapillary venules in lymph nodes: Observer variation and reproducibility in subjective estimation and morphometric measurements. To be published in *Acta Stereol.* 1984.
- Kraemer H. Extension of the kappa coefficient. *Biometrics* 1980; 36: 207-216.
- Lambourne A, Lederer H. Effects of observer variation in population screening for cervical carcinoma. *J Clin Path* 1973; 26: 564-569.
- Landis JR, Koch GG. The measurements of observer agreement for categorical data. *Biometrics* 1977; 33; 1: 158-174.
- Reyes MG, Tosch T, Childrey V, Penaranda LA, Noland CK. Reproducibility of stereologic analysis of light microscopic sections. In: Collan Y et al. (editors): *Stereology and morphometry in pathology.* Kuopio 1983. Also published as a special issue of *Acta Stereol* 2: 1983.
- Selkänaho K. Deriving coefficients of internal consistency of measurements: ICC and kappa. *Reports on Statistics, University of Jyväskylä* 1983; 12: 1-16.
- Sessions HA. Agreement and disagreement between pathologists in histological diagnosis. *Postgrad Med J* 1975; 51:685-689.
- Silcocks PBS. Measuring repeatability and validity of histological diagnosis - a brief review with some practical examples. *J Clin Pathol* 1983; 36: 1269-1275.
- Syrjänen KJ. The lymph nodes. Reactions to experimental and human tumors. *Exp Pathol* 1982; Suppl 8: 1-123.

ACKNOWLEDGMENTS

This study was supported by a grant from the North-Savo Regional Fund of the Finnish Cultural Foundation.