

REPRODUCIBILITY IN MORPHOMETRIC ASSESSMENT OF POSITIVE
STAINING FOR CEA IN OVARIAN TUMOURS

Veli-Matti Kosma, Yrjö Collan, Marja-Leena Aalto,
Anders Seppä, Markus Rautiainen, Kalevi Selkänaho*

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

ABSTRACT

Positive staining for carcinoembryonic antigen (CEA) in 30 ovarian tumours was assessed subjectively and morphometrically. The kappa coefficient (κ) and the intraclass correlation coefficient (ICC) were used to assess reproducibility. On subjective grading by 5 observers κ was 0.51 (90% confidence limits: 0.39-0.63) and by one observer (5 estimates) 0.96 (0.92-1). On morphometry the ICC of individual fields (5-person test) was 0.49-0.81 (0.26-0.93) and of standardized fields 0.64-0.78 (0.28-0.96), respectively. The ICC in 12 samples (5-person test, interobserver reproducibility) was 0.78-0.93 (0.31-0.99). In an intraobserver test (5 measurements) ICC was 0.87-0.95 (0.45-0.98). The results suggest that morphometric assessment is more reproducible than subjective assessment and that the human variations in counting account for most of the variation in morphometric reproducibility.

INTRODUCTION

Carcinoembryonic antigen (CEA) has been detected immunohistochemically in many ovarian tumours (Marchand et al., 1975, Rutanen et al., 1978), but its assessment has been mainly subjective and its reproducibility in diagnostic situation is suspect (Collan 1982). In the study by Aalto et al. (1982) the morphometric methods were used for evaluating positive CEA staining. The latter study revealed considerable biological and methodological variation in study settings planned for practical diagnostic purposes. So, in our study we set out to assess the reproducibility of CEA quantitation.

MATERIALS AND METHODS

A 5 micrometer thick section from each of 30 ovarian tumours were stained by the immunoperoxidase method (Aalto et al., 1982). First, each of the 30 ovarian tumours was considered a sample and total positivity was subjectively graded into four grades: negative (0), slight (1), moderate (2), and heavy (3) staining. The heavy staining (3) corresponded to the staining of colorectal adenocarcinoma. Subjective grading was further analysed in two experiments. In the first, 30 samples were graded once by 5 pathologists. In the second, the same samples were graded five times by one pathologist, who was experienced in the grading procedure. Second, one 5 micron thick section from each of 12 ovarian tumours were stained and assessed morphometrically (Aalto et al., 1982) once by 5 observers and five times by one observer, experienced in these measurements. Third, one single slide from one of tumours was analysed by 5 observers in two ways using a 25-point ocular grid (eye-piece graticule): A. Each observer selected 25 fields separately and randomly and made field and point scores for each field. Also the sample field and point scores were calculated. B. The section was sampled again with 25 square fields. The position of the fields was predetermined to give an even coverage of the whole sample. All five observers analysed these same 25 fields and the scores of individual fields were calculated as above. Also sample field and point scores were calculated. The kappa coefficient (Landis and Koch 1977, Kraemer 1980, Selkainaho 1983) and the intraclass correlation coefficient (ICC) (Cochran 1968, Selkainaho 1983) were calculated for subjective grades, scores of individual fields, and for sample field and point scores.

RESULTS

Reproducibility of the assessment of CEA staining analysed by the kappa coefficient and the intraclass correlation coefficient (ICC) is shown in Table 1.

DISCUSSION

The study shows that the reproducibility of morphometric assessment may vary considerably. In this our results parallel our earlier data on kidney (Collan et al., 1982) and the findings of Reyes et al. (1983) on muscle. The cause of variation in reproducibility (in terms of confidence limits) was clarified by pathologists studying 25 standard fields in one section. Because the images were identical there should have been no variation in reproducibility due to field selection or section

Table 1. Reproducibility of the assessment of CEA staining in terms of absolute values of kappa coefficient and intraclass correlation coefficient (ICC). The values present the reproducibility in subjective grading, in morphometry of 12 samples, in morphometry of randomly selected fields of one sample and in morphometry of standardized fields of the same sample.

<u>SUBJECTIVE GRADING</u>			
Interobserver reproducibility		Intraobserver reproducibility	
kappa		kappa	
0.51*** (0.39-0.63)		0.96***** (0.92-1)	
<u>MORPHOMETRIC MEASUREMENT</u>			
Interobserver reproducibility		Intraobserver reproducibility	
ICC		ICC	
Sample field score; epithelium	0.85*** (0.32-0.98)	0.95***** (0.87-0.98)	
Sample field score; mucin	0.93***** (0.51-0.99)	0.90***** (0.63-0.98)	
Sample point score; epithelium	0.78*** (0.37-0.94)	0.94***** (0.87-0.98)	
Sample point score; mucin	0.89***** (0.31-0.99)	0.87***** (0.45-0.98)	
<u>MORPHOMETRY; EFFECT OF STANDARDIZED FIELDS</u>			
Interobserver reproducibility		Interobserver reproducibility	
Random selection		Standardized selection	
ICC		ICC	
Field score; epithelium	0.49 ⁰ (0.27-0.68)	0.66**	(0.56-0.74)
Field score; mucin	0.49 ⁰ (0.26-0.69)	0.78***	(0.28-0.96)
Point score; epithelium	0.54* (0.41-0.66)	0.64**	(0.54-0.72)
Point score; mucin	0.81***** (0.54-0.93)	0.73***	(0.47-0.88)

⁰poor, * slight, ** fair, *** moderate, ***** substantial, ***** almost perfect degree of reproducibility.

thickness variation, and minimal variation in reproducibility if variation inherent to the method was a major cause of variation in reproducibility. However, the results did not show a dramatic decrease in relation to random fields. So, variation in counting (human factors) seems to be important. This need not be the case with ocular grids only but also applies other type of interpretation (Reyes et al., 1983). The reproducibility of the subjective evaluation of the total amount of positivity was smaller than that of morphometry. The interobserver reproducibility was much worse than intraobserver reproducibility. This suggests that biased estimates are common. These can probably be reduced only by uniform education of practitioners of morphometry.

REFERENCES

- Aalto M-L, Collan Y, Romppanen T, Lauslahti K. Morphometric approach to immunohistochemistry: Carcinoembryonic antigen (CEA) in ovarian tumours. *Acta Stereol* 1982; 1: 347-356, *Stereol* 82 Sheffield.
- Cochran WG. Errors of measurements in statistics. *Technometrics* 1968; 10: 637-666.
- Collan Y. Reproducibility, the neglected cornerstone of medical diagnosis. In: Collan Y, Romppanen T (editors): *Morphometry in Morphological Diagnosis*. Kuopio University Press, Kuopio 1982: 5-21.
- Collan Y, Romppanen T, Seppä A, Karhunen J, Jantunen E. Interobserver variation in morphometric kidney biopsy interpretation. In: Collan Y and Romppanen T (editors): *Morphometry in Morphological Diagnosis*. Kuopio University Press, Kuopio 1982: 155-161.
- Kraemer H. Extension of the kappa coefficient. *Biometrics* 1980; 36: 207-216.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33; 1: 158-174.
- Marchand A, Fenoglio EM, Pascal R, Richart RM, Bennett S. Carcinoembryonic antigen in human ovarian neoplasms. *Cancer Research* 1975; 35: 3807-3810.
- Reyes MG, Tosch T, Childrey V, Penaranda LA, Noland CK. Reproducibility of stereologic analysis of light microscopic sections. In: Collan Y et al. (editors): *Morphometry and stereology in pathology*. Kuopio 1983. *Acta Stereol* 2: 1983.
- Rutanen EM, Lindgren J, Sipponen P, Stenman UH, Saksela E, Seppälä M. Carcinoembryonic antigen in malignant and non-malignant gynecologic tumours. Circulating levels and tissue localization. *Cancer* 1978; 42: 581-590.
- Selkänaho K. Deriving coefficients of internal consistency of measurements: ICC and kappa. *Reports on Statistics, University of Jyväskylä* 1983; 12: 1-16.