

The reproducibility of nuclear morphometric measurements in invasive breast carcinoma*

Pauliina Kronqvist**, Teijo Kuopio and Yrjö Collan
Department of Pathology, University of Turku, Turku, Finland

Csaba Horvath
Semmelweis University, Budapest, Hungary

Ülle Tamm
University of Tartu, Tartu, Estonia

Received 16 January 1997

Revised 12 May 1997

Abstract. The intraobserver and interobserver reproducibility of computerized nuclear morphometry was determined in repeated measurements of 212 samples of invasive breast cancer. The influence of biological variation and the selection of the measurement area was also tested. Morphometrically determined mean nuclear profile area (Pearson's r 0.89, grading efficiency (GE) 0.95) and standard deviation (SD) of nuclear profile area (Pearson's r 0.84, GE 0.89) showed high reproducibility. In this respect, nuclear morphometry equals with other established methods of quantitative pathology and exceeds the results of subjective grading of nuclear atypia in invasive breast cancer. A training period of eight days was sufficient to produce clear improvement in consistency of nuclear morphometry results. By estimating the sources of variation it could be shown that the variation associated with the measurement procedure itself is small. Instead, sample associated variation is responsible for the majority of variation in the measurements (82.9% in mean nuclear profile area and 65.9% in SD of nuclear profile area). This study points out that when standardized methods are applied computerized morphometry is a reproducible and reliable method of assessing nuclear atypia in invasive breast cancer. For further improvement special emphasize should be put on sampling rules of selecting the microscope fields and measurement areas.

Keywords: Breast cancer, morphometry, nuclear profile area, reproducibility, nuclear pleomorphism, sources of variation

1. Introduction

Intuitively, quantitation is connected with the idea of high accuracy, reproducibility and reliability of the results. However, also quantitative methods involve sources of variation which influence the results and the conclusions. Sources of variation may make it difficult to adopt the method for use in clinical practice as a decision support in differential diagnostics and prognostically problematic cases.

*This study was supported by grants from the Finnish Cancer Research Foundation, the Cancer Foundation of South-West Finland, Ida Montin Foundation, and Turku University Hospital.

**Corresponding author: Dr P. Kronqvist, Department of Pathology, University of Turku, Kiinamylynkatu 10, SF-20520 Turku, Finland. Fax: +358 2 3337459.

Histological grading of breast cancer is widely used but it could be made more useful with better standardisation of the method [7]. As part of our effort of producing a quantitative grading system for breast cancer the purpose of this investigation is to examine the reproducibility of nuclear morphometry. Special emphasis is placed on intraobserver and interobserver variation of nuclear morphometry, on training of morphometric measurement technique, and on the selection of the measurement area.

2. Material and methods

2.1. Material

The study comprises 212 histologically verified cases of invasive breast cancer diagnosed at Turku University Hospital during the years 1990–1991 (Table 1). Of the samples 110 were fixed in buffered formalin and embedded in paraffin whereas 102 samples were first frozen and embedded in paraffin after frozen section diagnosis. Sections were cut at 5 μm and stained with haematoxylin and eosin. According to the original routine histopathologic diagnosis 174 cases were of ductal, 23 cases of lobular and 15 cases of special types of infiltrative breast cancer. Histological grades [6] were: grade 1, in 59 cases; grade 2, in 114 cases; and grade 3, in 39 cases.

2.2. Measurement of nuclear profile area

Measuring instrument. The nuclear profile area of the samples was measured using an image overlay drawing system run by the Prodit morphometry program (Prodit 3.1, Promis Inc, Almere, The Netherlands). The system includes a microscope, a personal computer (Compaq Deskpro 386/20e; Compaq Computer Corporation, Houston, TX, USA), a video camera (JVC TK-870U; JVC, Japan) and a digitizer board (PIP-512B video digitizer board; Matrox Electronic Systems, Dorval, Quebec, Canada). Digitized images of nuclear profiles were outlined on the monitor screen with a computer mouse [27]. At the beginning of each measuring session the system was calibrated with a micrometer slide. Measurements were performed with $\times 40$ objective magnification which when added to the $10\times$ video ocular and $2\times$ internal magnification resulted in an image of $\times 2500$ magnification on the monitor screen.

Sampling rule. The analysis of each sample began with selection of the *measurement area* at the clearly invasive border of the tumour in which the most cellular area was chosen for analysis. Necrotic and inflammatory areas were excluded. All distinguishable tumour cell nuclei in a microscopic *measurement field* were systematically selected starting from the upper-left corner of the measurement field. Altogether 6–15 adjacent fields were analysed until a total of 50 nuclei were measured.

Table 1

The samples examined by the observers 1 and 2 during the training period and during the analysis of interobserver variation. Profiles of 50 nuclei were measured from 5 μm paraffine sections. The measurement areas were selected by the trainer

Phase	Duration	Number of samples
Primary training	3 days	8
Final training	5 days	12
Analysis of interobserver variation	2 weeks	192
prefrozen material		102
nonfrozen material		110

2.3. Training of the observers

Observers 1 and 2 (CH and ÛT) were originally unfamiliar with the measuring method. The measuring method was described to the observers step by step whereafter they rehearsed the measuring procedure thoroughly. The trainer (PK) always selected the measurement area and marked it with ink. The measuring performance was under regular supervision by the trainer and conversations concerning the histological interpretation and measuring technique took place daily. Towards the end of the first training week (primary training phase) the observers worked together both measuring eight unselected samples twice. The results of the primary training phase were analyzed on the fifth day of the rehearsal period. At that time, means to reduce the influence of the sources of variation in the measurements were thoroughly discussed. In the next training phase of three days (final training phase) the observers worked independently. Applying the sampling rules earlier agreed on they carried out repeated measurements of additional 12 unselected samples. The results of the training phase were analysed at eight days of training.

2.4. Analysis of reproducibility of morphometry results

After the training period observers one and two measured nuclear profile areas of 192 samples to analyse the reproducibility associated with the morphometric measuring performance. The observers measured the nuclei from the same measuring areas selected and marked by the trainer. Measurement design of observers one and two is summarized in Table 1.

2.5. Analysis of the sources of variation

In a morphometric system the variation sources are intraobserver variation (V_m), interobserver variation (V_o) and inter-area variation (V_a) (Table 2) [10]. When variation is expressed as variances the figures are additive and the total variation in a histopathology laboratory (V_t) can be calculated

$$V_t = V_m + V_o + V_a.$$

For estimation of V_m , one pathologist (PK) measured 20 unselected samples three times. V_o was estimated after measurements of the same 20 samples by three observers. For estimation of V_a the three pathologists (YC, TK, PK) studied the same 20 cases and marked the measurement area of their choice with an ink mark encircling an area of approximately 3 mm in diameter on each slide according to the criteria presented earlier. The cases were assessed independently and the ink marks were always wiped off before the slides were passed on to the next pathologist. In each of the chosen areas one pathologist (PK) carried out the morphometric measurements.

The variances V_m , V_o and V_a were determined by estimating the average coefficient of variation ($CV = SD/\text{mean}$) of the measurements in each situation and then calculating the corresponding SD for

Table 2

Analysis of sources of variation in measuring nuclear profiles areas. A total of 20 samples were measured to estimate the size of interobserver (V_m), intraobserver (V_o), and inter-area (V_a) variation

Variation source	Number of observers	Number of measurement sessions
Intraobserver variation (V_m)	1	3
Interobserver variation (V_o)	3	1
Inter-area variation (V_a)*	1	3

*The measurement area was independently selected by three pathologists.

a theoretical situation in which the mean nuclear area was determined $40 \mu\text{m}^2$, and the SD of nuclear area was $10 \mu\text{m}^2$. Variances were squares of these SDs. The true interobserver variation of the system is calculated by subtraction of the intraobserver variation from the total variation after measurements by three different observers from the same fields. The true inter-area variation of the system is determined by subtraction of intraobserver variation from the total variation after measurements from three measurement areas selected by different pathologists. As a consequence, the true total variance (V_t) is the sum of the intraobserver variation, the true interobserver variation and the true inter-area variation. This approach allows us to determine the influence of each source of variation as a fraction of the total variation.

2.6. Statistical analysis

Pearson's correlation coefficients and Spearman's rank-order correlation coefficients were used to compare the measurements of observers one and two. Grading efficiency (GE) [11,13,14,28] was calculated to estimate the fraction of samples which can be correctly graded by the method under study. The GE's were determined from 2×2 , and 3×3 tables which estimate the results of two observers in classifying samples in two and three groups, respectively. In a similar approach, kappa coefficients were calculated to determine the relation of the observed agreement between the measurements to the expected at random agreement. Kappa values gave an estimate of the internal consistency of the method applied [26,31,38].

3. Results

The results of the measurements in the material of 192 breast cancer cases are outlined in Fig. 1 for mean nuclear profile area, and in Fig. 2 for SD of nuclear profile area. The averages of the results were identical in the measurements of both observers – mean nuclear profile area $41.9 \mu\text{m}^2$ (median $36.9 \mu\text{m}^2$) and standard deviation of nuclear profile area $12.9 \mu\text{m}^2$ (median $11.5 \mu\text{m}^2$). The correlation coefficients (Pearson's r and Spearman's ρ) are presented in Table 3. The results show a clear improvement in intra- and interobserver reproducibility during the training phase. The results classified into two and three groups are presented in Tables 4 and 5 in terms of GE's and kappa coefficients and demonstrate a corresponding trend. Figure 3 shows the distribution of GE's for mean nuclear profile area and Fig. 4 the distribution of SD of nuclear profile area at different cutpoints. The minimum GE was 0.93 for mean nuclear profile area (at the cutoff point of $34 \mu\text{m}^2$) and the minimum GE for the SD of nuclear profile area was 0.89 (at the cutoff of $10 \mu\text{m}^2$).

In Table 6 the variations are expressed as coefficients of variation (CV). The CV values were the lowest for intraobserver variations and by far the highest for inter-area variations. The true interobserver variation as expressed in variances is:

$$V_o = 7.44 - 4.25 = 3.19 \quad \text{for mean nuclear profile area, and}$$

$$V_o = 2.83 - 1.02 = 1.81 \quad \text{for SD deviation of nuclear profile area.}$$

Correspondingly, the true inter-area variation in terms of variances is:

$$V_a = 40.45 - 4.25 = 36.20 \quad \text{for mean nuclear profile area, and}$$

$$V_a = 4.45 - 1.02 = 3.43 \quad \text{for SD of nuclear profile area.}$$

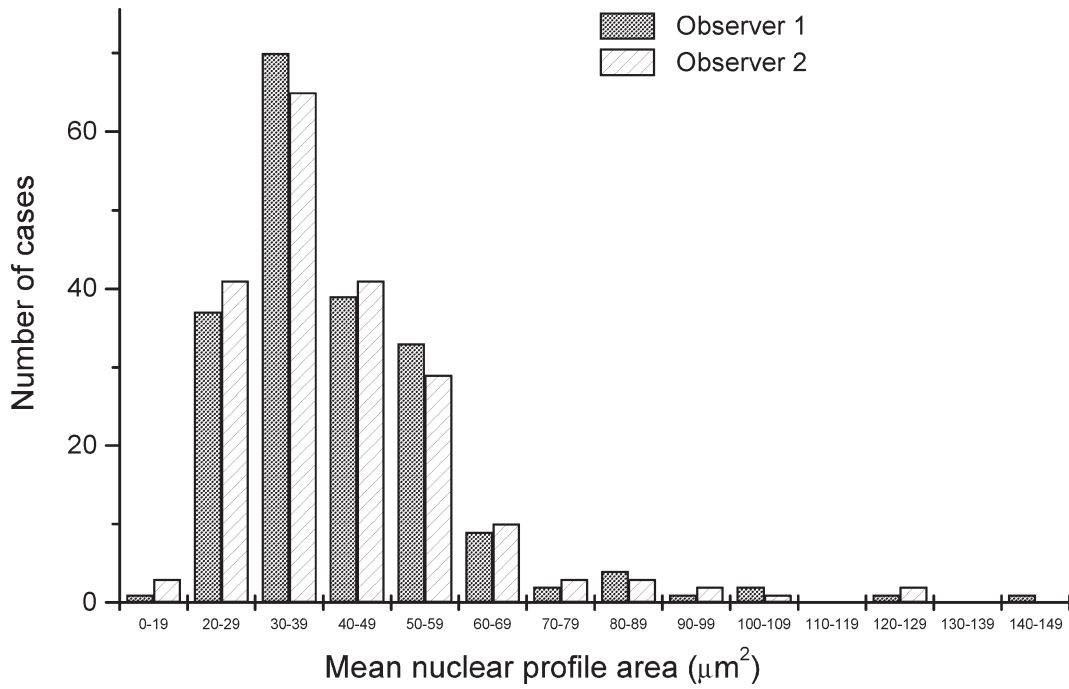


Fig. 1. Distribution of the mean nuclear profile area in the whole material of 192 breast cancer samples. The consistency of the measurements is excellent with the average of the mean nuclear profile areas $41.9 \mu\text{m}^2$ (median $36.9 \mu\text{m}^2$) for both observers.

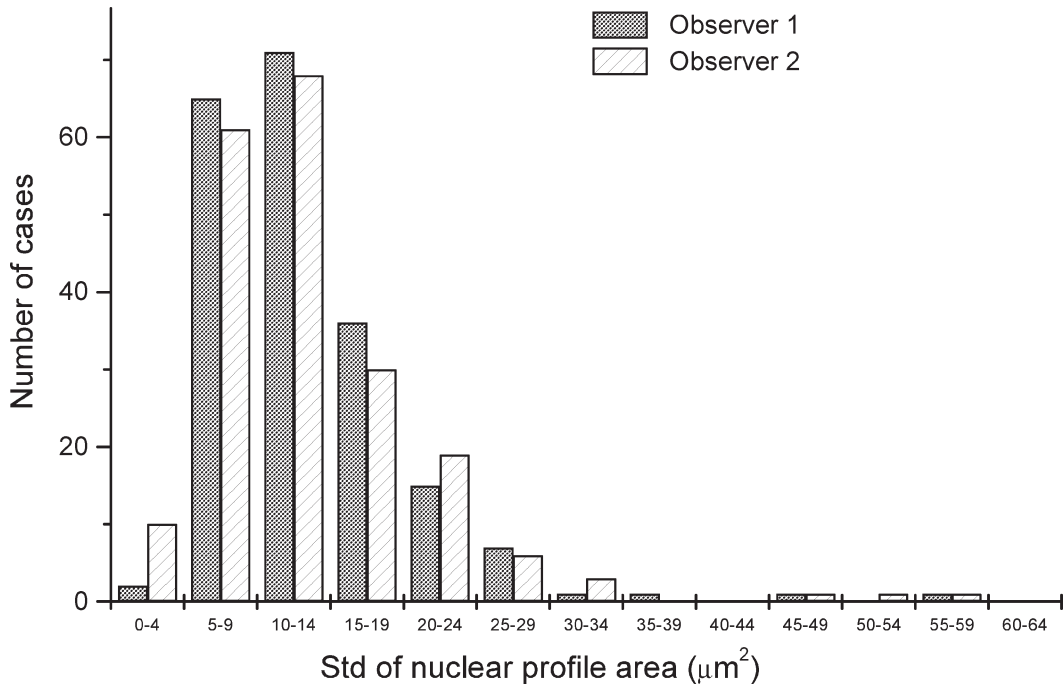


Fig. 2. Distribution of the standard deviation (SD) of nuclear profile area in the material of 192 samples of invasive breast cancer. Also the average SD was the same ($12.9 \mu\text{m}^2$, median $11.5 \mu\text{m}^2$) for both observers.

Table 3

Correlation coefficients (Pearson's r) and rank-order correlation coefficients (Spearman's ρ) of the morphometrically determined nuclear sizes (mean nuclear profile area; standard deviation of nuclear profile area in brackets). In the material of 192 breast cancer samples the two observers (Obs1 and 2) performed the measurements once. Before the measurement of the above cases, 20 breast cancer cases were used for training in two phases. The latter measurements were performed twice by each observer

	Single measurements by two observers (interobserver reproducibility)			Repeated measurements by one observer (intraobserver reproducibility)	
	r	ρ		r	ρ
192 breast cancer cases	0.89 (0.84)	0.88 (0.74)		–	–
20 breast cancer cases					
Primary training phase ($n = 8$)					
First measurement	0.58 (0.56)	0.57 (0.50)	Obs1	0.79 (0.73)	0.81 (0.79)
Second measurement	0.96 (0.78)	0.86 (0.83)	Obs2	0.66 (0.74)	0.55 (0.76)
Average of the above	0.77 (0.67)	0.71 (0.67)	Average	0.72 (0.74)	0.68 (0.77)
Final training phase ($n = 12$)					
First measurement	0.99 (0.98)	0.94 (0.53)	Obs1	0.97 (0.98)	0.80 (0.84)
Second measurement	0.99 (0.98)	0.94 (0.75)	Obs2	0.96 (0.96)	0.77 (0.60)
Average of the above	0.99 (0.98)	0.94 (0.64)	Average	0.96 (0.97)	0.79 (0.72)

Table 4

Grading efficiency of the morphometrically determined mean nuclear profile area (standard deviation of nuclear profile area in brackets). In the material of 192 breast cancer samples the two observers performed the measurements once. The thresholds for classifying the material into two groups were $35 \mu\text{m}^2$ for mean nuclear profile area and $10 \mu\text{m}^2$ for standard deviation of nuclear profile area. The corresponding thresholds for classifying the material into three groups were $30 \mu\text{m}^2$ and $40 \mu\text{m}^2$, and $7 \mu\text{m}^2$ and $13 \mu\text{m}^2$ for mean nuclear area and standard deviation of nuclear profile area, respectively. Before the measurement of the above cases, 20 breast cancer samples were used for training in two phases. The latter measurements were performed twice by each observer

	Single measurements by two observers (interobserver reproducibility)			Repeated measurements by one observer (intraobserver reproducibility)	
	r	ρ		r	ρ
192 breast cancer cases					
Two groups	0.95 (0.89)			–	–
Three groups	0.91 (0.84)			–	–
20 breast cancer cases					
Primary training phase ($n = 8$)					
First measurement	0.88 (0.75)		Obs1	0.75 (0.88)	
Second measurement	0.88 (0.81)		Obs2	0.88 (0.82)	
Average of the above	0.88 (0.78)		Average of the above	0.81 (0.84)	
Final training phase ($n = 12$)					
First measurement	0.92 (0.83)		Obs1	0.92 (0.88)	
Second measurement	0.96 (1.00)		Obs2	0.72 (0.92)	
Average of the above	0.94 (0.92)		Average of the above	0.82 (0.90)	

As a consequence, the total variance (V_t) is:

$$V_t = 4.25 + 3.19 + 36.20 = 43.64 \quad \text{for mean nuclear profile area, and}$$

$$V_t = 1.02 + 1.81 + 3.43 = 6.26 \quad \text{for SD of nuclear profile area.}$$

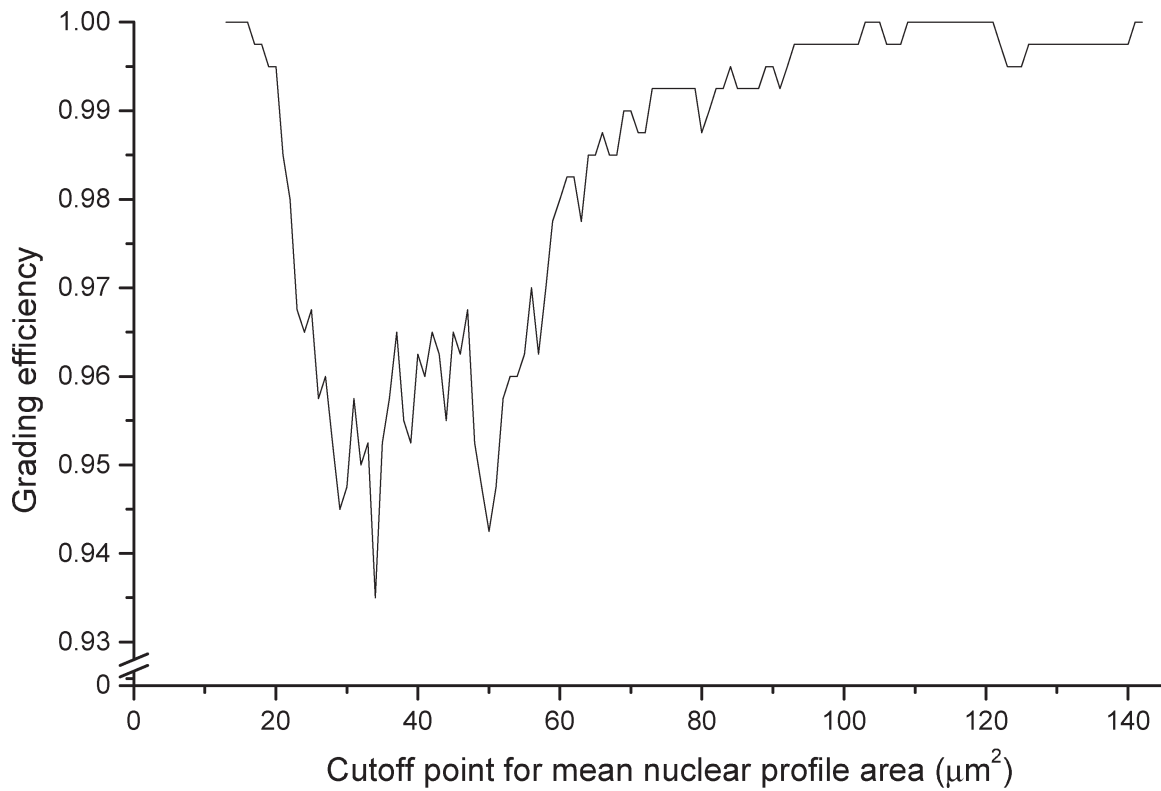


Fig. 3. The distribution of grading efficiencies (GE) for mean nuclear profile area at different cutpoints. The minimum GE is 0.93 at the cutoff of 34 μm^2 .

Table 5

Kappa coefficients of the morphometrically determined mean nuclear profile area (standard deviation of nuclear profile area in brackets). In the material of 192 breast cancer samples the two observers performed the measurements once. The thresholds for classifying the material into two groups were 35 μm^2 for mean nuclear profile area and 10 μm^2 for standard deviation of nuclear profile area. The corresponding thresholds for classifying the material into three groups were 30 μm^2 and 40 μm^2 , and 7 μm^2 and 13 μm^2 for mean nuclear area and standard deviation (SD) of nuclear profile area, respectively. Before the measurement of the above cases, 20 breast cancer samples were used for training in two phases. The latter measurements were performed twice by each observer

	Single measurements by two observers (interobserver reproducibility)		Repeated measurements by one observer (intraobserver reproducibility)
192 breast cancer cases			
Two groups	0.68 (0.53)		–
Three groups	0.75 (0.48)		–
20 breast cancer cases			
Primary training phase ($n = 8$)			
First measurement	0.53 (0.06)	Obs1	0.06 (0.52)
Second measurement	0.47 (0.14)	Obs2	0.53 (0.34)
Average of the above	0.50 (0.09)	Average of the above	0.30 (0.43)
Final training phase ($n = 12$)			
First measurement	0.68 (0.40)	Obs1	0.68 (0.53)
Second measurement	0.84 (1.00)	Obs2	0.67 (0.82)
Average of the above	0.76 (0.70)	Average of the above	0.67 (0.67)

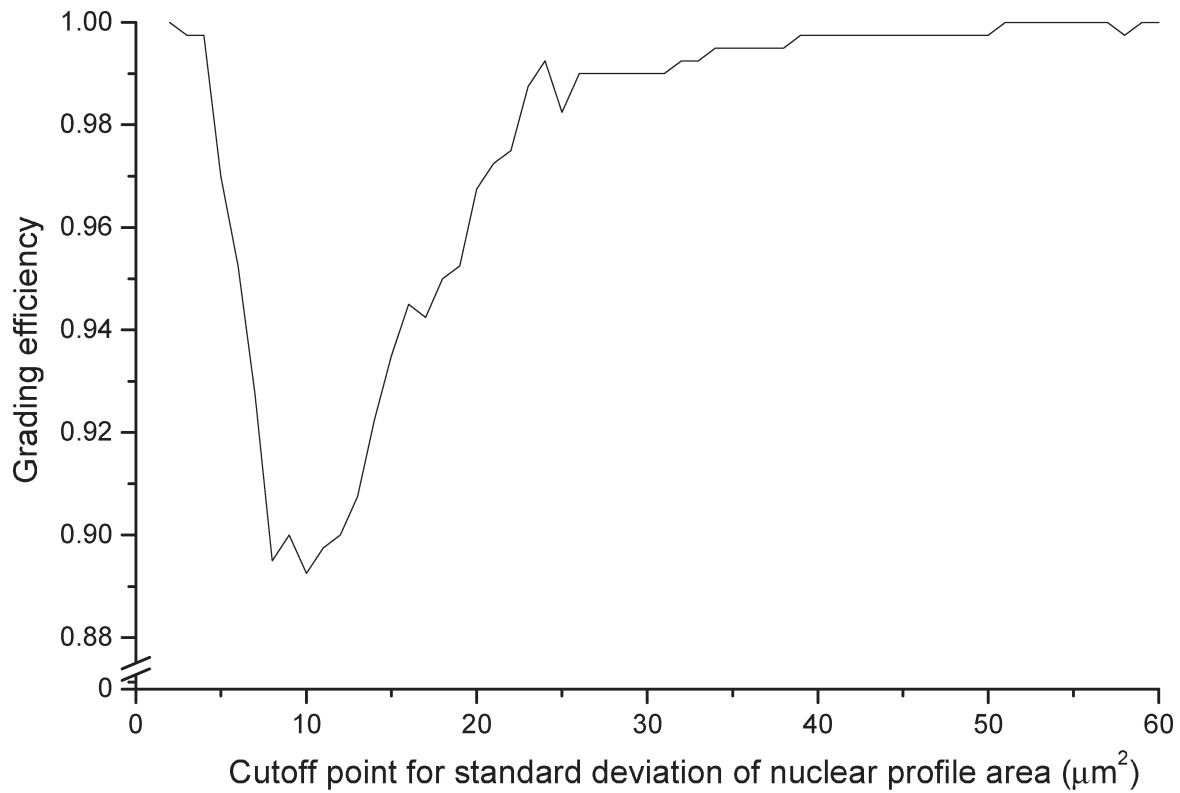


Fig. 4. The distribution of grading efficiencies (GE) for SD of nuclear profile area at different cutpoints. The minimum GE is 0.89 at the cutoff of $10 \mu\text{m}^2$.

Table 6

Variation of nuclear morphometric measurements by one observer in one measurement area, by one observer in three different measurement areas (selected by three pathologists), and by three measurements from the same area. This experiment was done to estimate the influence of field selection on the measurement results. $40 \mu\text{m}^2$ has been used to represent the theoretical average nuclear area, and $10 \mu\text{m}^2$ to represent the SD of nuclear area of all measurements

	Measurements from 1 area (intraobserver variation)		
	mean CV	area corrected SD	variance (SD2)
Mean nuclear profile area	0.052	2.063	4.25
SD of nuclear profile area	0.101	1.01	1.02
	3 measurements from 1 area (interobserver variation)		
	mean CV	area corrected SD	variance (SD2)
Mean nuclear profile area	0.061	2.73	7.44
SD of nuclear profile area	0.168	1.68	2.83
	1 measurement from 3 areas (inter-area variation)		
	mean CV	area corrected SD	variance (SD2)
Mean nuclear profile area	0.159	6.36	40.45
SD of nuclear profile area	0.211	2.11	4.45

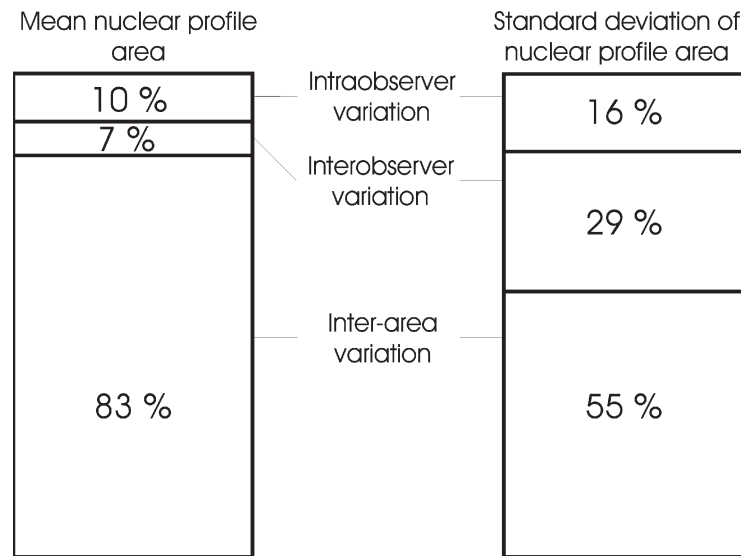


Fig. 5. The relations of the different sources of variation in the morphometric measuring method for mean nuclear profile area and for SD of nuclear profile area. The variation associated with the selected measurement areas and microscope fields is responsible for the largest fraction of the total variation. This reflects the biological variation within single samples.

The corresponding SD for a diagnostic situation in which all variation sources are present, are:

$$SD = \sqrt{43.64} = 6.61 \quad \text{for mean nuclear profile area, and}$$

$$SD = \sqrt{6.26} = 2.50 \quad \text{for SD of nuclear profile area.}$$

Finally, the fraction of total variation for mean nuclear profile area are:

$$V_m = 4.25/43.64 = 0.0974 = 9.7\%,$$

$$V_o = 3.19/43.64 = 0.0731 = 7.3\%,$$

$$V_a = 36.2/43.64 = 0.8295 = 83.0\%,$$

and for SD of nuclear profile area are:

$$V_m = 1.02/6.26 = 0.1629 = 16.3\%,$$

$$V_o = 1.81/6.26 = 0.2891 = 28.9\%,$$

$$V_a = 3.43/6.26 = 0.5479 = 54.8\% \text{ (Fig. 5).}$$

4. Conclusions

Barry and Sharkey [3] studied observer reproducibility of nuclear morphometry in 96 mammary ductal carcinoma samples embedded in plastic, sectioned at $1 \mu\text{m}$ and measured with $63\times$ oil-immersion objective. The correlation coefficients between measurements of two observers ranged from 0.840 to 0.910, and between repeated measurements of one observer from 0.889 to 0.915 (Pearson's r). In a

similar morphometric measurement setting, the results of nuclear morphometry in lymphocytic diffuse malignant lymphoma cells [8], however, showed a still better intraobserver consistency (Pearson's r) ranging from 0.977 to 0.995. Our results were well in line with the former study even though we used standard histological processing. This suggests that detailed measurements of individual nuclei do not markedly decrease variation in morphometry. On the other hand, the cell size in lymphocytic lymphoma is very uniform which explains the extremely high correlation observed. Concerning other methods of quantitative pathology, the reproducibility of the assessment of mitotic activity has been most thoroughly investigated [1,5]. Kujari with co-workers [28] studied the reproducibility of volume fraction-corrected mitotic index (M/V index, also called standardized mitotic index, SMI) [12] in 144 breast cancer specimens by four independent observers and two methods of analysing the epithelial fraction. The resulting correlation coefficients among methods ranged from 0.568 to 0.677 and between methods from 0.484 to 0.734. The corresponding mean interobserver GE varied between 0.90 and 0.93 (minimum 0.83). In another study on mitotic activity [19] the reproducibility of the assessment of MAI (mitotic activity index) in terms of Pearson's correlation coefficient ranged from 0.81 to 0.96. The latter results were based on mitotic counts in 13 pathology laboratories. Also, the inter-laboratory agreement of flow cytometric DNA-analysis in breast cancer is at the level of the present results on nuclear morphometry [2,4,20,21,29,30,32,33,37]. The conclusion is that morphometric measurements are accurate enough for creating a morphometric grading system, which – when biologically relevant – could compete with other well established methods of quantitative pathology.

In a study by Delides and co-workers [17] a perfect agreement among six pathologists of subjectively assessed final grade was observed in only 14.5% of the cases. In Scarff–Bloom–Richardson gradings of breast cancer an agreement among six observers was found for 74.3% of all specimens [15,16,23–25,35–37]. Data on the reproducibility of the subjective assessment of nuclear pleomorphism in breast carcinoma have not been available. In our unpublished material Pearson's r of subjective nuclear gradings of invasive breast cancer ranged from 0.56 to 0.68 (mean 0.62) in repeated gradings by one observer. The corresponding mean GE and K were 0.86 and 0.58, respectively. Between observers, Pearson's r ranged from 0.44 to 0.59 (mean 0.53), and GE and kappa were 0.77 and 0.39, respectively. The results of the present paper favour the morphometric approach over subjective assessment to reach accuracy of grading.

In the early phases of starting morphometry, training clearly improved the reproducibility. The variation associated with the SD of nuclear area improved the most. By the end of the rehearsal, however, the assessment of the nuclear area and nuclear area variation were found almost equally reproducible. During the training period interobserver reproducibility was almost systematically higher than intraobserver reproducibility. In the primary training phase this can be explained by the fact that this part of rehearsal was teamwork with the observers discussing the possible applications of the measuring criteria in each sample. In part, the relatively low intraobserver reproducibilities were caused by the observers constantly correcting their measuring techniques to produce between themselves as uniform measurement results as possible. In the results, it may also seem confusing that the improvement of the reproducibilities of the results does not continue throughout the whole experiment but decreases while measuring the 192 cases. This finding may partly be explained by the biologic variation of the analysed feature and partly with increasing intraobserver variance in a larger material [3,8]. The fact that the reproducibility of results did not markedly improve during the latter part of the rehearsal also suggests that the training period of eight days can be considered sufficient.

Collan et al. [9] and Van Diest et al. [18] summarize the sources of variation in the nuclear morphometric method which involve factors related to the sample as well as tissue processing and

instrumentation. According to our results, intraobserver and interobserver variation both concerning mean nuclear area and SD of nuclear area were low ranging from 7.3% to 28.9%. The inter-area variation, in turn, was responsible for the majority of the total variation associated with the measuring method (83.0% in mean nuclear profile area and 54.8% in SD of nuclear area). In light of these figures, choosing the measurement areas was the most important factor influencing the reproducibility of the measurements. The consistency of the selection of the measurement areas in this study can be considered good. In fact, in 16 out of the 20 samples the three pathologists actually chose and marked the very same areas of the specimen for measurement and also in the rest of the samples the selected areas were located along the same invasive border of the sample. Thus the main part of the measurement associated variation does not seem to be due to the selection of the measurement area but instead is associated with the inconsistency between the microscopic fields analysed in the marked measurement area. The main conclusion is that the biological variation present within the sample is responsible for the majority of the total variation. This suggests that the reproducibility of morphometrically determined nuclear size cannot markedly be improved with more efforts in standardisation of the measuring method. Rather, stress should be put on sampling methods associated with choosing the measurement area and consistent selection of nuclei from the chosen area. Sampling rules including systematic sampling from many fields over the whole section could result in more stable results and higher range of nuclear size variation associated features. For this, many alternatives are available [22] including the application of stereology [34], but the disadvantage of laborious implementations is obvious especially to diagnostic pathologists. The development of measurement support programs may soon make these approaches more attractive.

References

- [1] J.P. Baak, P.J. van Diest, A.T. Ariens, M.W. van Beek, S.M. Bellot, J. Fijnheer, L.H. van Gorp, W.S. Kwee, J. Los and H.C. Peterse, Multicenter Morphometric Mammary Carcinoma Project (MMMCP). A nationwide prospective study on reproducibility and prognostic power of routine quantitative assessment in the Netherlands, *Path. Res. Pract.* **185** (1989), 664–670.
- [2] B. Baldetorp, P.O. Bendahl, M. Fernö, K. Alanen, U. Delle, U. Falkmer, B. Hansson-Aggesjö, T. Höckenström, A. Lindgren, L. Mossberg, S. Nordling, H. Sigurdsson, O. Stål and T. Visakorpi, Reproducibility in DNA flow cytometric analysis of breast cancer: Comparison of 12 laboratories' results for 67 sample homogenates, *Cytometry* **22** (1995), 115–127.
- [3] J.D. Barry and F.E. Sharkey, Observers reproducibility during computer-assisted planimetric measurements of nuclear features, *Hum. Pathol.* **16** (1985), 225–227.
- [4] E. Bergers, P.J. van Diest and J.P. Baak, Reproducibility of semi-automated cell cycle analysis of flow cytometric DNA-histograms of fresh breast cancer material, *Anal. Cell. Pathol.* **8** (1995), 1–13 (abstr.).
- [5] S. Biesterfeld, I. Noll, E. Noll, D. Wohltmann and A. Böcking, Mitotic frequency as a prognostic factor in breast cancer, *Hum. Pathol.* **26** (1995), 47–52.
- [6] H.J.G. Bloom and W.W. Richardson, Histological grading and prognosis in breast cancer, *Br. J. Cancer* **11** (1957), 359–377.
- [7] Breast cancer Prognostic factors – the search goes on [News], *Journal of the National Cancer Institute* **86**(7) (1994), 480–482.
- [8] K.W. Chan, K.Y. Chiu, K.H. Fu and J.M.L. Ling, Observer variability in microcomputer-assisted morphometric study of nuclear parameters, *Pathology* **19** (1987), 407–409.
- [9] Y. Collan, R. Montironi, G.M. Mariuzzi, T. Torkkeli, F. Marinelli, E. Pesonen, G. Collina, V.-M. Kosma, E. Jantunen and O. Kosunen, Observer variation in interactive computerized morphometry, *Appl. Pathol.* **4** (1986), 9–14.
- [10] Y. Collan, T. Torkkeli, V.-M. Kosma, E. Pesonen, O. Kosunen, E. Jantunen, G.M. Mariuzzi, R. Montironi, F. Marinelli and G. Collina, Sampling in diagnostic morphometry: the influence of variation sources, *Path. Res. Pract.* **182** (1987), 401–406.
- [11] Y. Collan, General principles of grading lesions in diagnostic histopathology, *Path. Res. Pract.* **185** (1989), 539–543.

- [12] Y. Collan, T. Kuopio, J.P.A. Baak, R. Becker, W.V. Bogomoletz, M. Deverell, P. van Diest, C. van Galen, K. Gilchrist, A. Jared, V.-M. Kosma, H. Kujari, P. Luzi, G.M. Mariuzzi, E. Matze, R. Montironi, M. Scarpelli, D. Sierra, S. Sisti, S. Toikkanen, P. Tosi, W.F. Whimster and E. Wisse, Standardized mitotic counts in breast cancer. Evaluation of the method, *Path. Res. Pract.* **192** (1996), 931–941.
- [13] Y. Collan and E. Pesonen, Principles of morphometric grading, *Gegenbaurs Morphol. Jahrbuch* **135** (1989), 13–17.
- [14] Y. Collan, P. Klemi, O. Kallioniemi, H. Joensuu, S. Nordling and M. Eskelinen, Significance of variation in DNA flow cytometric analyses from paraffin-embedded breast cancers. Evaluation of the grading efficiency of ploidy determination, DNA index, and S-phase fraction, *Path. Res. Pract.* **188** (1992), 581–586.
- [15] L.W. Dalton, D.L. Page and W.D. Dupont, Histologic grading of breast carcinoma. A reproducibility study, *Cancer* **73** (1994), 2765–2770.
- [16] B.W. Davis, R.D. Gelber, A. Goldhisch, W.H. Hartmann, G.W. Locher, R. Reed, R. Golouh, J. Säve-Söderbergh, L. Holloway, I. Russel and C.M. Rudenstam, Prognostic significance of tumor grade in clinical trials of adjuvant therapy for breast cancer with axillary lymph node metastasis, *Cancer* **58** (1986), 2662–2672.
- [17] G.S. Delides, G. Garas, G. Georgouli, D. Jiortziotis, J. Lecca, T. Liva and J. Elemenoglou, Intralaboratory variations in the grading of breast carcinoma, *Arch. Pathol. Lab. Med.* **106** (1982), 126–128.
- [18] P.J. van Diest, A.W.M. Smeulders, F.B.J.M. Thunnissen and J.P.A. Baak, Cytomorphometry. A methodological study of preparation techniques, selection methods and sample sizes, *Anal. Quant. Cytol. Histol.* **11** (1989), 225–231.
- [19] P.J.A. van Diest, J.P. Baak, P. Matze-Cok, E.C.M. Wisse-Brekelmans, C.M. van Galen, P.H.J. Kurver, S.M. Bellot, J. Fijnheer, L.H.M. van Gorp, W.S. Kwee, J. Los, J.L. Peterse, H.M. Ruitenbergh, R.F.M. Schapers, M.E.I. Schipper, J.G. Somsen, A.W.P.M. Willig and A.Th. Ariens, Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the Multicenter Morphometric Mammary Carcinoma Project (MMMCP), *Hum. Pathol.* **23** (1992), 603–607.
- [20] L.G. Dressler, L. Seamer, M.A. Owens, G.M. Clark and W.L. McGuire, Evaluation of a modelling system for S-phase estimation in breast cancer by flow cytometry, *Cancer Res.* **47** (1987), 5294–5302.
- [21] M. Fernö, B. Baldetorp, S.-B. Ewers, I. Idvall, H. Olsson, H. Sigurdsson and D. Killander, One or multiple samplings for flow cytometric DNA-analyses in breast cancer – prognostic implications?, *Cytometry* **13** (1992), 241–249.
- [22] J.C. Fleege, P.J. van Diest and J.P.A. Baak, Reliability of quantitative pathological assessments, standards, and quality control, in: *Manual of Quantitative Pathology in Cancer Diagnosis and Prognosis*, J.P.A. Baak, ed., Springer, Berlin, 1991.
- [23] K.W. Gilchrist, L. Kalish, V.E. Gould, S. Hirschl, J.E. Imbriglia, W.M. Levy, A.S. Patchefsky, D.W. Penner, J. Pickren, J.A. Roth, R.A. Schinella, I.S. Schwarz and J.E. Wheeler, Interobserver reproducibility of histopathological features in stage II breast cancer. An ECOG study, *Breast Cancer Res. Treat.* **5** (1985), 3–10 (abstr).
- [24] D.S. Hopton, J. Thorogood, A.D. Clayden and D. MacKinnon, Observer variation in histological grading of breast cancer, *Eur. J. Surg. Oncol.* **15** (1989), 21–23.
- [25] L.P. Howell, R. Gandour-Edwards and D. O’Sullivan, Application of the Scarff–Bloom–Richardson tumour grading system to fine-needle aspirates of the breast, *Am. J. Clin. Pathol.* **101** (1994), 262–265.
- [26] H. Kraemer, Extension of the kappa coefficient, *Biometrics* **36** (1980), 207–216.
- [27] P. Kronqvist, Y. Collan, T. Kuopio and H. Kujari, Nuclear morphometry in breast cancer: the influence of sampling rules and freezing of samples, *Mod. Path.* **8** (1995), 187–192.
- [28] H.P. Kujari, Y.U.I. Collan and N.B. Atkin, Use of the mitotic counts for the prognosis and grading of breast cancer. Method evaluation study, *Path. Res. Pract.* **190** (1994), 593–599.
- [29] T.E. Kute, B. Gregory, J. Gallenshaw, M. Hopkins, D. Buss and D. Case, How reproducible are flow cytometry data from paraffin-embedded blocks?, *Cytometry* **9** (1988), 494–498.
- [30] T.E. Kute, H.B. Muss, M.R. Cooper, L.D. Case, D. Buss, V. Stanley, B. Gregory, J. Gallenshaw and K. Booher, The use of flow cytometry for the prognosis of stage II adjuvant treated breast cancer patients, *Cancer* **66** (1990), 1810–1816.
- [31] J.R. Landis and G.G. Koch, The measurement of observer agreement for categorical data, *Biometrics* **33** (1977), 159–174.
- [32] B. Risberg, B. Baldetorp, M. Fernö, B. Hansson, L.L. Mossberg, G. Roos, O. Stål, J.T. Thornthwaite and L. Tibell, Inter-institutional reproducibility of flow cytometric DNA-analysis in breast carcinomas, *Anal. Cell. Pathol.* **6** (1994), 23–36 (abstr).
- [33] R. Silvestrini, Quality controls for evaluation of the S-phase fraction by flow cytometry: a multicentric study. The SICCAB Group for Quality Control of Cell Kinetic Determinations, *Cytometry* **18** (1994), 11–16.
- [34] F.B. Sörensen, Biology of the disease. Quantitative analysis of nuclear size for objective nuclear grading: a review with emphasis on new, unbiased stereological methods, *Lab. Invest.* **66** (1992), 4–23.
- [35] B. Stenkvist, S. Westman-Naeser, J. Vegelius, J. Holmquist, B. Nordlin, E. Bengtson and O. Eriksson, Analysis of reproducibility of subjective grading systems for breast carcinoma, *J. Clin. Pathol.* **32** (1979), 979–985.

- [36] F. Theissig, K.D. Kurze, G. Haroske and W. Mayer, Histological grading of breast cancer. Interobserver reproducibility and prognostic significance, *Pathol. Res. Pract.* **186** (1990), 732–736.
- [37] F. Theissig, V. Dimmer, G. Haroske, K.D. Kurze and W. Meyer, Use of nuclear image cytometry, histopathologic grading and DNA cytometry to make breast cancer prognosis more objective, *Anal. Cell. Pathol.* **3** (1991), 351–360 (abstr.).
- [38] H.R. Wulff, *Pålideligheden af Deskriptive Data. Rationel Klinik*, Grundlaget for diagnostiske og terapeutiske beslutninger, Copenhagen, 1981, pp. 48–54.