Virchows Arch (1998) 433:135–143

ORIGINAL ARTICLE

Oleksiy Tsybrovskyy · Inna Vassilenko
Sebastian Mannweiler · Martin Klimpfinger

Multivariate karyometric approach in differential diagnosis of follicular thyroid neoplasms

A study of 31 cases

Received: 1 August 1997 / Accepted: 23 March 1998

Abstract A retrospective analysis of 19 follicular adenomas, 12 minimally invasive follicular carcinomas and 3 widely invasive follicular carcinomas of the thyroid was performed on 5-µm-thick Feulgen-stained paraffin sections by means of a semiautomatic system for picture analysis. The major aim was to assess the potential of multiparameter karyometry for separation of the first two tumour types. Sixteen planimetric and densitometric features were defined in each case on 200–300 randomly selected nuclei and processed by a number of uni- and multivariate statistical methods. Despite predominantly significant ANOVA results a substantial overlap between tumour groups limited the practical usefulness of any karyometric feature alone. Factor and cluster analyses indicated independence of planimetric and densitometric parameters from each other, which was of crucial importance in finding an optimal subset of variables for discriminant analysis. The classification rule derived from the latter procedure was checked by the "jack-knife" method, by classification of 3 widely invasive cancers and by hierarchical tumour clustering. Sensitivity and specificity of the model for detection of malignancy were 100% and 94.7%, respectively. A multivariate karyometric approach, when applied correctly, can be a useful tool for differentiation between follicular adenomas and minimally invasive follicular carcinomas of the thyroid.

O. Tsybrovskyy[1] (✉) · I. Vassilenko
Department of Anatomic Pathology,
State Medical University of Donetsk, Donetsk, Ukraine

S. Mannweiler
Department of Pathology, University of Graz,
School of Medicine, Graz, Austria

M. Klimpfinger
Department of Pathology and Bacteriology,
Kaiser-Franz-Josef-Hospital, Vienna, Austria

Mailing address:
[1] c/o H. Denk, MD, FRC Path, Department of Pathology,
Karl Franzens University of Graz, Auenbruggerplatz 25,
A-8036 Graz, Austria
Fax: (+43)-316-38-43-29

Key words Thyroid neoplasms · Diagnosis, differential · Image processing, computer assisted · Karyometry · Multivariate analysis

Introduction

One of the major problems in thyroid surgical pathology is the differential diagnosis of follicular adenomas and well-differentiated, minimally invasive follicular carcinomas [15]. According to the current rules of histological typing of thyroid tumours proposed by WHO [19], the usual criteria for malignancy, like cellular atypia and mitotic activity, are not helpful, and the only reliable indicator of malignancy is capsular penetration and/or blood vessel invasion. Indeed, in most cases cytological examination of the fine needle aspiration material fails to differentiate follicular carcinomas from adenomas, especially atypical ones [34, 39]. This uncertainty results in a large number of non-conclusive cytology reports or, in attempts to produce more concrete judgments, leads to a high percentage of either false-positive or false-negative diagnoses, depending on the strategy chosen by the pathologist [39]. In suggestive cases other methods, including intraoperative study of frozen sections, are required [40]. However, this last-named procedure has the same drawbacks, because of its low sensitivity for thyroid carcinomas in general (53%) and for well-differentiated follicular carcinomas in particular (0%) [35]. Even in paraffin sections careful examination of numerous slides is necessary to confirm the presence or absence of invasion [15, 22, 34]. Thus, the distinction between these two types of thyroid tumours is a challenging task for the pathologist. One possible way to improve diagnostic accuracy is to take a quantitative karyometric approach. This method already has a fairly long history in thyroid pathology, being regarded as very promising by some [1, 23, 24] and useless by other authors [25, 28, 30, 42, 43]. Most investigations support an "intermediate" standpoint: that any statistically significant difference revealed by karyometry is usually in-

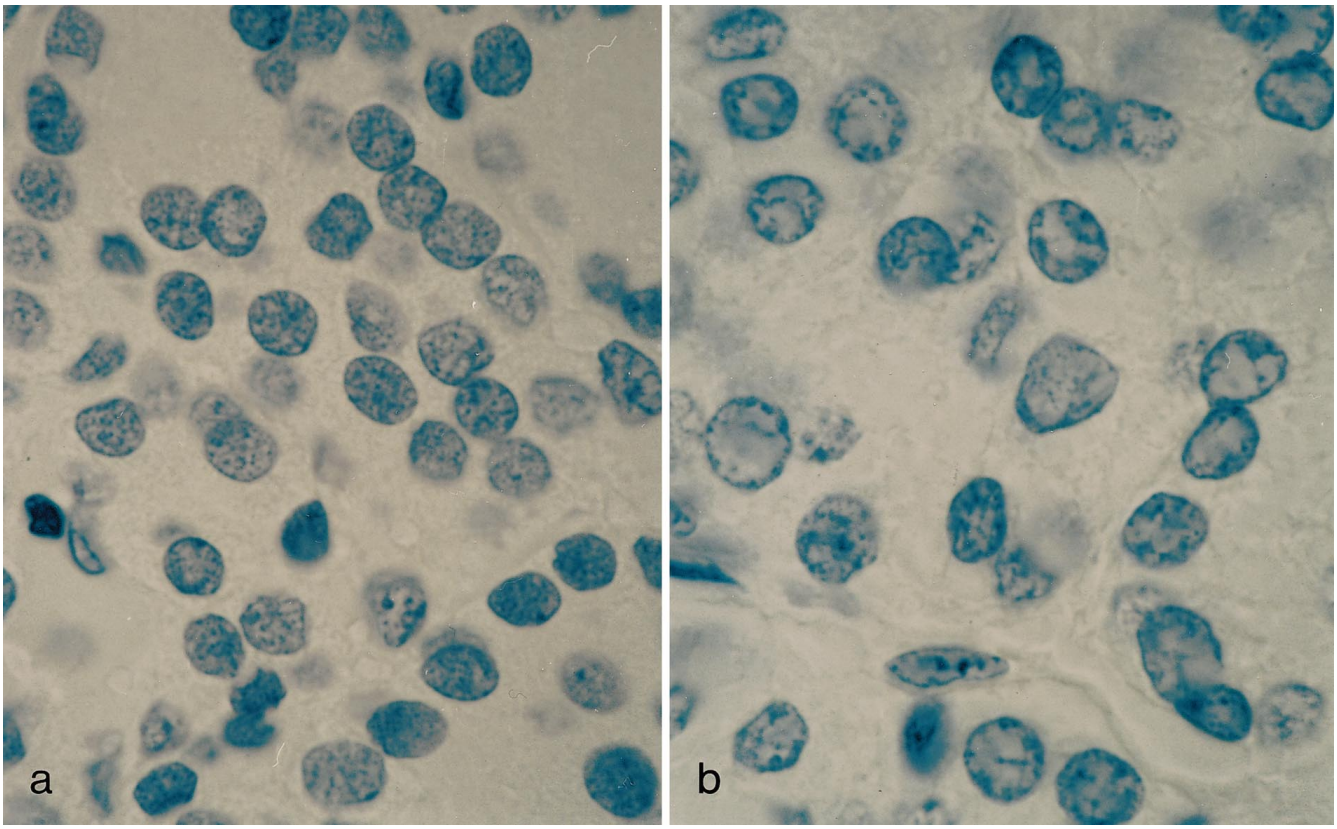
sufficient per se for correct classification of individual cases because of the wide overlap of tumour groups [11, 16, 17, 29, 36, 38, 44, 47]. However, the performance of this method is greatly improved when a multivariate statistical approach is adopted. In fact, it has become almost the rule in quantitative cytology that not only one but several features have to be measured for each cell in a sample. Under these conditions, multivariate statistical methods are much more appropriate and powerful than comparison of group means by Student's *t*-test or analysis of variance based on Chi-square distribution [29]. Baak and associates [1] were the first to show the great advantage of stepwise multiple regression analysis in cytology-based classification of thyroid lesions, but their ideas were not elaborated further until a few years ago. By using linear discriminant analysis, many authors created highly accurate classification models based on a combination of diverse karyometric [8, 26, 27, 32] or nucleolar [31] features. Another multivariate method, principal component analysis, has been successfully applied mainly in theoretical rather than practical settings [7, 37]. Nevertheless, pessimistic conclusions have also been drawn after using this technique [4, 12]. It should be stressed that few authors have used a multivariate approach in the differential diagnosis of follicular thyroid tumours – and then mostly with contradictory results [4, 8, 12, 31, 32] – and only one has applied it to histological material [32].

The major aim of this study was to look for possible ways of separating follicular adenomas and minimally invasive follicular carcinomas of the thyroid gland in paraffin sections with the aid of karyometry followed by a multivariate statistical processing of the data.

Materials and methods

In all, 19 follicular thyroid adenomas and 15 follicular thyroid carcinomas (archival material) were analysed. The case selection criteria were based on the WHO classification [19]. Only adenomas with trabecular, microfollicular or normofollicular patterns of growth and so-called atypical adenomas were included. Of the cancers, 12 were minimally invasive and 3 widely invasive in type. Macrofollicular adenomas, multinodular goitres with nodular hyperplasia, follicular variant of papillary thyroid carcinoma and purely oxyphilic neoplasms were excluded. All surgically removed specimens were fixed in buffered 10% formalin and embedded in paraplast. In every case, new 5-µm-thick Feulgen stained sections were made (Fig. 1 a, b). To ensure an adequate and unbiased sampling, the following rules were followed. The fields of vision were selected randomly from a whole tumour region according to a so-called raster method, which is both highly efficient and reproducible [13, 14]. The minimally required number of nuclei for a sample to be representative was determined by preliminary measurements as described by Collan et al. [6] and was approximately 100; Fleege et al. have shown that equal sample sizes are sufficient for the raster method in the majority of cases [13]. To avoid any doubt we measured at least twice as many (200–300) nuclei per slide. The focus was adjusted separately for every neoplastic cell in a field, because any deviations from correct focus greatly affect the precision of the measurements [3, 10]. Only nuclei with clearly defined borders and without overlap were eligible, and only 3–5

Fig. 1 **a** Nuclei from a solid-trabecular thyroid adenoma. **b** Nuclei from a minimally invasive follicular thyroid carcinoma. Feulgen stain, original magnification ×1000



of them were actually measured in a frame, as recommended by Fleege et al. [14]. Nuclei with obvious signs of mitosis, apoptosis, or other degenerative changes were excluded.

Karyometric features were evaluated by means of a semiautomatic system for picture analysis. The hardware configuration included a 3CCD colour video camera, model DXC-930P (SONY, Japan) mounted on an Axioplan light microscope (ZEISS, Germany) and an image processing unit (CIRES, KONTRON, Germany) based on a powerful microcomputer with Pentium processor. Nuclear images were read through a Neofluar 1.3/100 objective with oil immersion, digitized by a frame grabber at 752×582 spatial resolution and 256 levels of grey and projected on to a TV monitor at ×1800 final magnification. For the analysis we used the red image channel, since it is complementary to the blue colour and thus allows the best discrimination of different picture elements [5]. Further image processing was carried out by using CIRES software, version 3.0a. Nuclear contours were traced either automatically or interactively, depending on the degree of separation between a nucleus and the background. The following parameters were measured [5]:

Nucleus area (NA)
Mean optical density within a nucleus (MOD)
Standard deviation of optical density within a nucleus (SDOD)
Skewness in the frequency distribution of optical density within a nucleus (SkewOD)
Excess in the frequency distribution of optical density within a nucleus (ExcOD)
Minimal optical density within a nucleus (MinOD)
Maximal optical density within a nucleus (MaxOD)
Integrated optical density within a nucleus (IOD)

Thus, one planimetric feature (NA) characterized the size of neoplastic nuclei, whereas 6 densitometric parameters described several different chromatin properties, such as degree of condensation (MOD, MinOD, MaxOD), heterogeneity (SDOD) or distribution within a nucleus (SkewOD, ExcOD) [7]; the last feature, IOD, reflected the general amount of the absorbing material in a given nucleus [5]. The system was calibrated with a micrometric scale to obtain the NA values in square micrometre. All remaining parameters were measured in arbitrary units related to a reference image; calibration of illumination and correction for shading and local background were used each time to avoid any artificial changes in object brightness [3]. The data obtained were stored on the hard disk and then exported to Excel spreadsheets. At this stage, the values of all parameters were averaged separately for each patient; moreover, standard deviations, as measures of internuclear variability, were also calculated. The final data set, therefore, contained the doubled (16) number of columns (variables) and the number of rows already corresponding to the tumours studied (19 for adenomas and 15 for carcinomas). All further uni- and multivariate statistical procedures were carried out on this data set exclusively by means of SPSS for Windows software (SPSS, Chicago, Ill.), version 6.1.3a. With respect to the following steps, however, it was unwise to ignore the heterogeneity of the carcinoma group (that is, the mixture of cancers with minimal and massive invasion of the stroma). Indeed, the widely invasive carcinomas might display rather distinct karyometric features compared with the minimally invasive ones, and moreover, they usually do not present any diagnostic difficulties for the pathologists. However, it was also irrational to split the carcinoma group, since there were too few massively invasive carcinomas. Thus, for the sake of subsequent analysis, we decided not to include the widely invasive cancers until the last stages and to focus mainly on the differences between the adenomas and the carcinomas with a minimal invasion of the stroma.

Results

All variables were checked for deviation from normality by the one-sample nonparametric Kolmogorov-Smirnov test. In no case could the null hypothesis be rejected (significance levels ranged from 0.20 to 0.99 for both groups), so that we could assume a Gaussian distribution of the parameters. On this basis multivariate normality was also assumed, because at present it cannot be tested directly in any other way [2, 9, 33]. The assumption of normality permitted the application of parametric methods in the next steps. One-way analysis of variance (ANOVA) was used for univariate comparisons of the group means. For all variables, covariance matrices were assumed to be equal (*P*-values in the Levene test ranged from 0.06 to 0.98). ANOVA revealed statistically significant differences between follicular adenomas and carcinomas for all parameters measured except for IOD Mean (Table 1). Two variables – MinOD Mean and SkewOD Mean – had higher *F*-scores than the others, suggesting a

Table 1 Karyometric features of follicular thyroid neoplasms evaluated at group means, and ANOVA-test results (*NA* nucleus area, *MOD* mean optical density within a nucleus, *SDOD* standard optical density within a nucleus, *ExcOD* excess in the frequency distribution of optical density within a nucleus, *IOD*, integrated optical density within a nucleus, *MinOD MaxOD* minimum, maximum optical density within a nucleus, *StDev* standard deviation)

Variable no.	Parameter[a]	Follicular adenomas		Follicular carcinomas[b]		One-way ANOVA	
		Mean	SEM	Mean	SEM	F-value	Significance
1	NA Mean	31.65	1.81	47.35	1.97	32.3	<0.00005
2	NA StDev	7.32	0.72	13.93	1.22	24.9	<0.00005
3	MOD	0.35	0.02	0.24	0.01	26.3	<0.00005
4	OD StDev	0.08	0.004	0.06	0.004	11.2	0.002
5	SDOD Mean	0.10	0.006	0.08	0.004	5.9	0.02
6	SDOD StDev	0.03	0.002	0.02	0.001	13.8	<0.00005
7	SkewOD Mean	0.48	0.06	–0.09	0.05	46.8	<0.00005
8	SkewOD StDev	0.47	0.01	0.38	0.02	17.2	<0.00005
9	ExcOD Mean	0.12	0.07	–0.24	0.03	14.1	<0.00005
10	ExcOD StDev	0.68	0.04	0.46	0.02	21.3	<0.00005
11	MinOD Mean	0.13	0.005	0.08	0.004	50.0	<0.00005
12	MinOD StDev	0.05	0.004	0.04	0.003	7.4	0.01
13	MaxOD Mean	0.67	0.03	0.54	0.03	7.9	0.009
14	MaxOD StDev	0.16	0.01	0.12	0.008	5.7	0.04
15	IOD Mean	1128.3	75.8	1150.9	37.45	0.04	0.84
16	IOD StDev	267.5	25.2	361.5	16.61	5.5	0.03

[a] Values of parameters are given for NA in square micrometers and for all others in arbitrary units
[b] The carcinoma group included only minimally invasive cancers

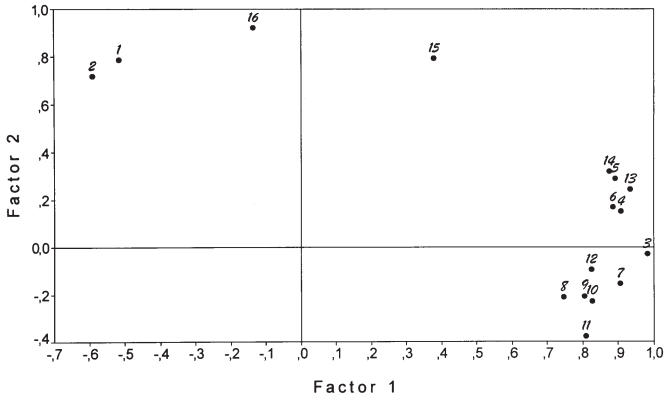


Fig. 2 Plot of correlations (factor loadings) between karyometric parameters and first two principal components extracted by factor analysis, after Quartimax rotation. Variables are coded ordinally by numbers from 1 to 16 in accordance with Table 1

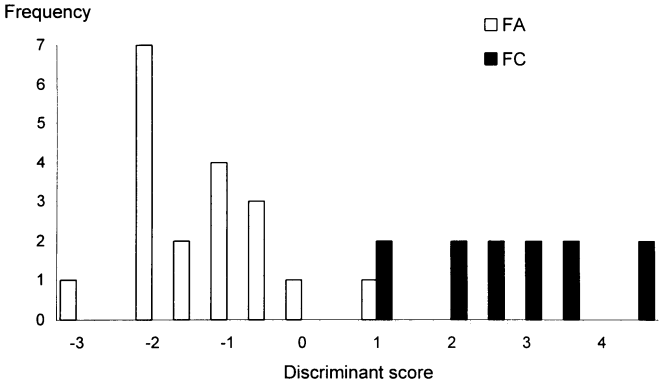


Fig. 3 Frequency histogram of discriminant scores in the two tumour groups (*FA* follicular adenoma, *FC* minimally invasive follicular carcinoma

leading role for them in tumour distinction. It should be emphasized that, despite the highly significant ANOVA results, none of the variables alone was sufficient for classification of individual cases (for example, by using cut-off values) because of a consistent overlap of tumour groups.

To investigate possible interactions among variables, we applied a number of multivariate statistical methods.

1. Factor analysis. First, all principal components having eigenvalues greater than 1 were extracted. Tests for sampling adequacy (Kaiser-Meyer-Olkin measure 0.73; Barlett's test of sphericity $P<0.0001$) and large communalities of the variables (ranging from 0.73 to 0.97) indicated the relevance and high efficiency of the analysis [2, 33]. For the simplest interpretation of both factors and variables, orthogonal Quartimax rotation of the structure matrix was chosen [33]. As a result, the program extracted 3 underlying principal components, which explained 88.8% of the total variance (61.9%, 19.8% and 7.1% by the first, second and third factors, respectively). The contribution of the third factor was relatively small (7.1%), and even after rotation all variables had their highest loadings on the two first principal components. The final structure matrix was, therefore, recalculated and plotted in two dimensions only, without taking the third factor into account, as shown in Fig. 2. Obviously, the entire set of karyometric parameters falls into two distinct categories. The first one contains all densitometric features (nos. 3–14), which are strongly correlated with factor 1. The four remaining variables having the highest loadings on factor 2 (NA Mean, NA StDev, IOD Mean, IOD StDev, nos. 1, 2, 15 and 16, respectively) form the second category. It should be pointed out that the same subdivision was also obtainable with other methods of orthogonal rotation (Varimax, Equamax) as well as without rotation at all (data not shown), but in either case factors remain uncorrelated with each other. We decided, therefore, to take into account the presence of these two independent parameter categories for further analyses. As will be shown below, such approach turned out to be of great importance in the next steps.

2. Discriminant analysis. To build a highly stable classification model, the number of variables should be restricted according to the size of the smallest group [2], that is to say 5 or 6 at most in the present study. Hence, the forward stepwise selection of variables based on minimization of Wilks' lambda was adopted. The value of F was ≥3.84 for entry (approx. $P=0.05$) and ≤2.71 for removal (approx. $P=0.1$). We did not use the significance levels as selection criteria, since for some reason the P-values computed by SPSS at each step are not exact [33]. The chosen strategy, despite its substantial advantages, is not ideal, and its performance can be rather poor when the general number of variables is large [9]. To overcome this drawback, we explored not only the whole data set but also each of the two parameter categories revealed by factor analysis. Then we compiled all variables selected from both categories and compared the results with the previous ones. When all variables were analysed simultaneously, only three densitometric parameters (OD Mean, SkewOD Mean and MinOD Mean) were selected. Analysis of the first parameter category gave the same results. From the second category NA Mean and IOD Mean were chosen. Finally, combination of all five variables resulted in the greatest decrease of the Wilks' lambda, which was statistically significant compared with any of the foregoing variants. Moreover, F-scores of the parameters exceeded the critical values for removal specified above. Thus, the canonical discriminant function for further analysis was computed on the basis of these five variables. Its Wilks' lambda was 0.193, which corresponded to a significance level of $P<0.0001$. In other words, the null hypothesis can be rejected. To observe the degree of separation between the tumours, discriminant scores of all cases were calculated and plotted as shown in Fig. 3. In spite of the obvious difference between the groups a slight overlap could be seen: that is, 1 adenoma was found in the area occupied by carcinomas.

Table 2 Predicted groups and posterior probabilities obtained in discriminant analysis by using two different methods of model validation. Widely invasive cancers were held out while computing the model. *FA* follicular adenoma, *FC* follicular carcinoma: *m.i.* minimally invasive; *w.i.* widely invasive; ** wrongly classified.

No.	Actual group	Resubstitution method		Jack-knife method	
		Predicted group	Probability	Predicted group	Probability
1	FA	FA	1	FA	1
2	FA	FA	1	FA	1
3	FA	FA	1	FA	1
4	FA	FA	1	FA	0.93
5	FA	FA	0.97	FA	0.96
6	FA	FA	0.89	FA	0.87
7	FA	FC**	0.79	FC**	0.96
8	FA	FA	1	FA	1
9	FA	FA	1	FA	1
10	FA	FA	0.99	FA	0.99
11	FA	FA	1	FA	1
12	FA	FA	1	FA	1
13	FA	FA	1	FA	1
14	FA	FA	1	FA	1
15	FA	FA	1	FA	1
16	FA	FA	1	FA	0.99
17	FA	FA	1	FA	1
18	FA	FA	1	FA	1
19	FA	FA	1	FA	1
20	$FC_{m.i.}$	FC	1	FC	1
21	$FC_{m.i.}$	FC	1	FC	1
22	$FC_{m.i.}$	FC	1	FC	1
23	$FC_{m.i.}$	FC	1	FC	1
24	$FC_{m.i.}$	FC	0.916	FC	0.877
25	$FC_{m.i.}$	FC	0.9964	FC	0.9912
26	$FC_{m.i.}$	FC	0.9994	FC	0.9992
27	$FC_{m.i.}$	FC	0.9038	FC	0.8751
28	$FC_{m.i.}$	FC	0.9992	FC	0.9984
29	$FC_{m.i.}$	FC	1	FC	1
30	$FC_{m.i.}$	FC	1	FC	1
31	$FC_{m.i.}$	FC	1	FC	1
1	$FC_{w.i.}$	FC	0.987	–	–
2	$FC_{w.i.}$	FC	1	–	–
3	$FC_{w.i.}$	FC	0.962	–	–

The fitness of the model was estimated by case reclassification, as described below. According to Box's M test, the group covariance matrices were not equal (P=0.0002). The posterior probabilities were therefore calculated on separate-groups covariance matrices. After that, the program assigned every case to the one of the two tumour groups for which the posterior probability was larger (Bayes' rule). Prior probabilities and costs of misclassification were assumed to be equal for both groups. To avoid an overoptimistic misclassification estimate, the final validation of the model was performed by using the jack-knife (leaving-one-out) method. It involves leaving out each of the cases in turn, calculating the discriminant function based on the remaining n-1 cases, and then classifying the left-out case [33]. Such a procedure, in contrast to simple case resubstitution, supplies unbiased results and has to be preferred to any other technique when sample sizes are relatively small [2, 9]. However, as shown in Table 2, the two validation

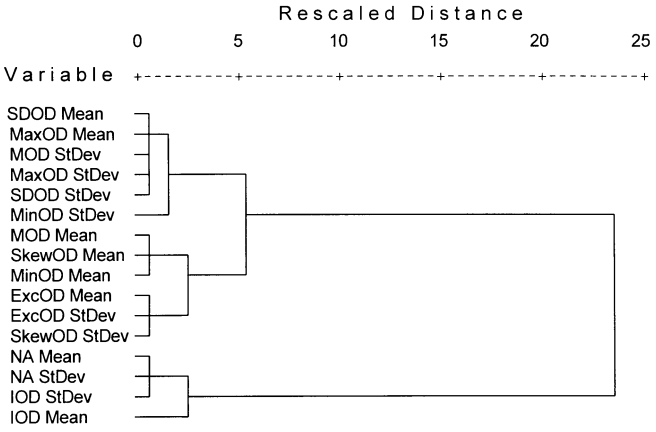


Fig. 4 Horizontal dendrogram representing steps in hierarchical clustering of karyometric parameters. The acutal distances at which separate clusters were joined up are proportionally rescaled to a range from 0 to 25 so that the ratio of the distances between steps is preserved

schemes in our example produced practically equal posterior probabilities and misclassification rates. One follicular adenoma (no. 7) was assigned by the model to the carcinoma group, whereas all other cases were classified correctly. It corresponds to an overall accuracy of 96.8%. Specificity and sensitivity of the method for malignancy detection were also calculated as described by Jasani and Schmid [20] and constituted 94.7% and 100%, respectively. It is worth mentioning that the same misclassification rates were produced by the two models based on separate parameter categories; however, posterior probabilities under these circumstances were far less convincing than those in Table 2, particularly when only planimetric features were used.

After the model was computed and validated, it was interesting to see how it would treat the widely invasive carcinomas, which had been withdrawn from the analysis. As shown in Table 2, all these 3 tumours were regarded by the model as malignant, and the probabilities of the correct diagnosis were very high.

3. Cluster analysis. We applied hierarchical clustering as an alternative to factor and discriminant analyses. That is, both cases and variables were grouped. In each version Ward's method for clustering was adopted, since it has some substantial advantages over all others and produces usually the most plausible classification schemes [2]. Squared Euclidean distance or Pearson correlation (in absolute values) was specified as a measure of dissimilarity between cases or variables, respectively. To equalize effects of differently scaled variables, all values in the data set were standardized to a range of 0–1 [33]. For case grouping, only the features selected in stepwise discriminant analysis were used. Steps in matching the variables are shown in Fig. 4. After agglomeration of all clusters with pairwise distances less than 6 rescaled units, there are only 2 clusters left, which correspond exactly to the parameter categories revealed by factor analysis. Again, these two groups of variables appear to be

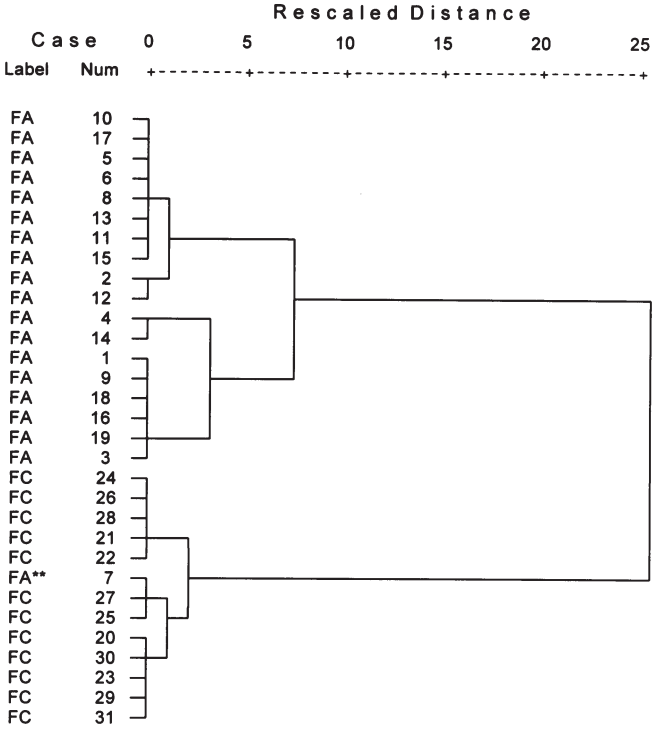


Fig. 5 Horizontal dendrogram representing steps in hierarchical clustering of follicular thyroid neoplasms on the base of five karyometric parameters (NA mean, MOD mean, SkewOD mean, MinOD mean und IOD mean). Case numbers correspond to those in Table 2 (** wrongly grouped). For further explanation see also Fig. 4

quite distinct from each other, since the rescaled distance between them is much larger than that of the next subdivision (24 and 6 units respectively; see Fig. 4). The summary of the case clustering can be interpreted in a similar way (Fig. 5). The two main clusters are identical to the groups predicted in discriminant analysis, including the number of the misclassified case (cf. Table 2). Furthermore, the tumour groups are also well separated and can be clearly defined in the greater part of the multidimensional space (rescaled distances from 7 to 25 units; see Fig. 5).

Discussion

According to current WHO recommendations, the diagnosis of minimally invasive follicular carcinoma should be based entirely on the demonstration of vascular and/or capsular invasion [19]. Such restriction of the diagnostic criteria is not satisfactory, for at least two reasons. The diagnosis of follicular adenoma becomes shaky without representative sampling of the tumour capsule, which is particularly frequent in consultation cases [15], and for the same reason an intraoperative examination of frozen sections is insensitive for minimally invasive follicular carcinoma [35]. The other point is that even in serial paraffin sections unequivocal recognition of true capsular invasion can sometimes be difficult. For example, penetration of the inner half of the capsule or appearance of tumour islands embedded within the capsule does not count as true invasion according to the common definitions [19, 34]. However, Schröder et al. point out that these islands sometimes represent real tumour invasion, although it is usually not possible to differentiate them histologically from islands of non-invasive adenoma tissue that have been entrapped within the capsule owing to capsular enfolding or capsular fibrosis [41]. Furthermore, both these authors [41] and Kahn and Perzin [22] reported development of metastatic disease in patients who had infiltration (but not complete disruption) of the capsule – the capsule being sectioned in toto – as the only histological proof of malignancy. Similarly, evidence of angioinvasion can be questionable [22]. Such uncertainty results in wide interobserver variability, which has been well documented by Greenbaum et al. [18]. Thus, the difficulties inherent in clear separation of these two tumour types seems to be a stumbling block in thyroid pathology. It is especially true when it is taken into account that immunohistochemistry, electron microscopy, and flow cytometric DNA measurements all fail to be helpful in this respect [20, 21, 41, 43, 45, 46].

We present a possible way of improving diagnostic precision by using the multivariate karyometric approach. The method combines high precision and reproducibility of karyometry with the power of multivariate statistics [29].

One-way analysis of variance revealed significant differences between follicular adenomas and minimally invasive follicular carcinomas on all parameters measured except for IOD Mean (see Table 1). Similar findings are reported by most other authors [7, 16, 17, 23–25, 32, 36, 38, 44, 47]. As for IOD Mean, this parameter is usually a starting point in calculation of DNA indices for ploidy assessments [5]. We did not perform such an analysis, because histological preparations used by us did not satisfy the recently accepted standards for DNA measurements [3]. Nonetheless, the non-significant ANOVA results for IOD Mean indirectly confirm the equality of ploidy levels in both types of follicular thyroid tumours [21, 30, 34, 42, 43, 45]. It should further be mentioned that some investigators who used a karyometric approach did not observe any statistically significant differences between follicular adenomas and carcinomas at all [28, 30, 42, 43]. The only obvious explanation for such disagreement is the small size of the samples, especially in the carcinoma group. Indeed, nowhere (including the present study) does the number of malignant follicular tumours exceed 27 (and no more than 10 with minimal invasion), which is a small number for statistical analysis. This shortcoming is not easy to overcome, however, because minimally invasive follicular carcinomas are rare [34]. As a result, there is usually an enhanced probability of a type II error in karyometry (failure to detect significant differences), which explains the aforementioned discrepancies. It deserves special emphasis that the matter in question is the *statistical*, and not the *practical* significance of the findings. That is to say that any

hypothesis testing simply answers the question as to *whether* the diagnostic groups are different, and gives little idea about *how* different they are and *how useful* this difference is in practice. Unfortunately, both the data in the literature [4, 24, 25, 36, 38] and our own experience indicate that even highly significant testing results in karyometry are still accompanied by an inevitable overlap between patient groups, which severely limits the diagnostic usefulness of any karyometric feature alone.

In contrast to univariate statistics, multivariate processing of the data provided us with much more information. Taking into account the results of factor analysis, we divided the whole set of variables into two independent categories. This separation was later confirmed by cluster analysis (see Figs. 2, 4). The first parameter category covered all densitometric features, which correlated with each other very closely. Thus, our initial surmise about different meanings of these variables failed to hold; they are probably all equally dependent on certain changes in the chromatin structure, and can be regarded as a united group of descriptive nuclear features [10]. The second parameter category included the characteristics related to the nuclear size, i.e. NA Mean, NA StDev, IOD Mean and IOD StDev. The presence of the last two variables here is justified, since the IOD value is calculated as the sum of optical densities of each pixel engrossed by a given nucleus on a TV screen [5]. As a consequence, this parameter depends about equally on both nuclear area and optical density. Figure 2 clearly demonstrates this relationship: IOD Mean (no. 15) lies practically in the middle between planimetric (1, 2) and densitometric (3–14) features. Yet, it is referred to the second category owing to a much higher loading on factor 2 (0.8) than factor 1 (0.38; see Fig. 2).

Thus, densitometric and planimetric properties of neoplastic nuclei proved to be independent of each other. Furthermore, their relative contribution to the discrimination between benign and malignant conditions was not the same. As described in the Results, factor 1 accounted for most (61.9%) of the total variance, only 19.8% being explained by factor 2. That is probably why only densitometric features were selected by the first run of stepwise discriminant analysis. When regarded separately, the first parameter category yielded more confident classification results than the second. We can conclude, therefore, that densitometric properties of neoplastic nuclei are superior to dimensional characteristics with respect to their discriminatory power. Unfortunately, it is impossible to correlate this finding directly with any reference in the literature, because to the best of our knowledge, nobody has used factor analysis previously to investigate possible interactions among different karyometric features. However, practically all authors who performed simultaneous measurements of several distinct (geometric, densitometric, textural) nuclear parameters came to analogous conclusions [16, 24, 26, 27, 37].

Nevertheless, we agree with Liautaud-Roger et al. [27] that planimetric features, in spite of their secondary position, have a high discriminatory potential. In this respect our results are remarkably similar to those obtained by Nafe et al. [32], even with regard to the absolute values of the parameters measured. As a result, we noted that the combination of both densitometric and planimetric features was significantly better than any of the categories alone. In practice, this superiority expressed itself not in lower misclassification rates but rather in a higher assurance of posterior probabilities. We, therefore, consider the two parameter categories as complementary, which is of crucial importance in finding an optimal subset of variables for discriminant analysis. We also want to emphasize that IOD Mean, while of no value in accordance with ANOVA results, contributed significantly to the discrimination between the tumour groups when used together with other 4 selected variables. This fact demonstrates once more the advantage of using multivariate statistical methods in a case when several measured parameters can somehow be correlated with each other.

Similarity of misclassification rates and posterior probabilities supplied by the two validation schemes (see Table 2) indicates the adequacy of the case-to-variable ratio in the analysis and high stability of the model [9]. This conclusion is also confirmed by completely identical case grouping obtained in cluster analysis (Fig. 5). Moreover, the cases of widely invasive carcinoma were classified by the model as definite malignancies (see Table 2), so we can indeed expect a good performance of the model in a prospective study. In addition, a relative uniformity of karyometric differences between follicular adenomas and carcinomas can be assumed regardless of how widely the malignancies invade the stroma. The latter surmise requires further verification, since only a few widely invasive cancers were involved in the current study.

It is striking that during all classification procedures performed, one follicular adenoma (no. 7) was always wrongly assigned by the program to the carcinoma group. This tumour had been resected in a 57-year-old woman and appeared grossly as a completely encapsulated solitary thyroid nodule with maximum diameter of 40 mm. At microscopical examination the tumour showed predominantly normofollicular pattern of growth. The capsula was of uneven thickness, and neoplastic cells exhibited conspicuous variability in nuclear size. A large number of gigantic nuclei strongly suggested the presence of aneuploidy. In addition, tumour cells focally underwent oxyphilic transformation. Serial sections of the entire tumour mass did not reveal any evidence of capsular or vascular invasion. The patient was alive and free of disease within 2 years after the operation. Hence, we found no reasons to change the initial diagnosis. Nevertheless, some authors believe that DNA aneuploidy may be a self-sufficient criterion for follicular thyroid cancer, even in the absence of obvious invasion [18]. Further reasoning in this direction will lead us to the dispute on whether preinvasive carcinoma or carcinoma in situ exists and how justifiable a sharp separation of benign and malignant follicular tumours is in general

[18, 30, 42, 43]; however, this topic is beyond the scope of the current paper. At present, we are inclined to conform to the generally accepted beliefs in this respect [19, 34] and attribute the wrongly classified case to the imperfect nature of our design rather than to biological heterogeneity of follicular thyroid neoplasms.

Hence, the classification rule obtained had an overall efficiency of 96.8%. We believe, however, that the true performance of a model should be assessed more precisely in a similar way to the rules adopted in immunocytochemistry for the quality estimation of primary antibody reagents [20], i.e. in terms of sensitivity and specificity for malignancy detection. Unlike the overall accuracy, which can be very misleading when one of the groups is much smaller than the other [33], these two indices always permit an objective assessment of the final output. In our study sensitivity of the model was 100% and its specificity 94.7%. These values are among the best described in the literature, although the actual misclassification rates are difficult to compare owing to the distinct validation procedures applied by other authors. The first multivariate statistical model (stepwise regression analysis) in the history of thyroid morphometry created by Baak et al. [1] was based on two planimetric features (nuclear/cytoplasmic size ratio and cell size) and showed very high (up to 100%) sensitivity and specificity for detection of malignancy. However, these encouraging results were substantially discredited by work of Luck et al. [28], who found it impossible to perform cytoplasmic measurements in fine needle aspirates because of the absence of definite cytoplasmic boundaries and also did not find any statistically significant differences between nuclear characteristics of the follicular thyroid tumours [28]. Nafe et al. [32] computed their discriminant function on a learning set composed of 13 follicular adenomas and 8 follicular carcinomas and then applied a testing set (4 follicular adenomas and 5 well-differentiated follicular carcinomas) to examine it. One case of adenoma was falsely classified, which means 75% specificity and 100% sensitivity. However, the hold-out method used by these authors has a substantial disadvantage: when the sample size is small, a considerable part of the information contained in the testing set is lost, and validation results are usually too pessimistic [2, 9, 33]. If we take into account that reclassification of the learning set in the work mentioned [32] was absolutely correct, then the specificity increases to 94.1%, which is already very close to our results. In a number of other papers the sensitivity and specificity values for follicular malignancies were 90.9% and 100% [8], 89% and 100% [12], and 96.2% and 93.8% [31], respectively, although some positive bias in these rates is present owing to the character of the validation procedure used [2, 9, 33].

It should be mentioned that not all attempts of this kind were so successful. For instance, the level of discrimination between microfollicular adenomas and follicular carcinomas obtained by Salmon et al. [36] was only 58% after stepwise discriminant analysis based on 15 morphometric, densitometric and textural parameters. Moreover, Cavallari et al. [4] demonstrated complete inability of such analysis to distinguish between benign and malignant follicular tumours in fine needle aspiration smears. We would like to point out, however, that only a few planimetric features were measured in the latter study [4].

What can be seen from this short literature review is a trend towards generation of highly accurate mathematical models for malignancy detection in follicular thyroid neoplasms, although there is some inconsistency of results obtained by different authors. These discrepancies are not surprising, because in every particular case a large number of different factors, such as type of pathological material used, types and subtypes of tumours selected for investigation, staining techniques, sampling modes, family of parameters measured, statistical methods adopted, to name but a few, make up a unique combination, which in the end produces results that are difficult to compare. From our experience, an appropriate choice of diverse options within a given multivariate statistical procedure proved to have a great impact on the contents and correctness of the final output; this is why we have described our data processing in such detail in section Results. Thus, it would be very desirable to standardize the above-mentioned determinants [6], as has recently been done, for instance, in DNA image cytometry [3], or at least to specify all the crucial settings more precisely.

In conclusion, the application of multiparameter karyometry in thyroid surgical pathology has not yet given unequivocal results, although a strong positive tendency in this field is evident. On the basis of the data obtained in our current study we believe that the method, when applied correctly, can become a very useful tool in differential diagnosis between follicular adenomas and minimally invasive follicular carcinomas of the thyroid gland.

Acknowledgments We are grateful to Dr. Andrea Berghold, Institute for Medical Information Technology, Statistics and Documentation, University of Graz, for her kind assistance. This work was supported by grants from the Austrian Federal Ministry of Science and Research and from the Institute of Pathology, Karl-Franzens-University of Graz, Austria. We would like also to express our particular gratitude to Prof. Dr. Helmut Denk, Department of Pathology, Karl Franzens University of Graz, Austria, for putting excellent working facilities at our disposal and also for general supervision on the project and substantial correction of the manuscript.

References

1. Baak JPA, Kurver PHJ, Boon ME (1982) Computer-aided application of quantitative microscopy in diagnostic pathology. Pathol Annu 17:287–306
2. Backhaus K, Erichson B, Plinke W, Weiber R (1996) Multivariate Analysemethoden. Springer, Berlin Heidelberg New York
3. Böcking A, Giroud F, Reith A (1995) Consensus report of the European Society for Analytical Cellular Pathology task force on standardization of diagnostic DNA image cytometry. Anal Quant Cytol Histol 17:1–6

4. Cavallari V, Maiorana A, La Rosa GL, Maiorana MC, Scimone S, Fano RA (1989) Morphometric studies on fine-needle aspirates from follicular proliferative lesions of the thyroid. Pathologica 81:411–446
5. CIRES (1994) Cell image retrieval and evaluation system . Kontron Elektronik GmbH, Munich
6. Collan Y, Torkkeli T, Kosma VM, Pesonen E, Kosunen O, Jantunen E (1987) Sampling in diagnostic morphometry: the influence of variation sources. Pathol Res Pract 182:401–406
7. Collin F, Salmon I, Rahier I, Pasteels JL, Heimann R, Kiss R (1991) Quantitative nuclear cell image analysis of thyroid tumors from archival material. Hum Pathol 22:191–196
8. Crissman JD, Drozdowicz S, Johnson C, Kini SR (1991) Fine needle aspiration diagnosis of hyperplastic and neoplastic follicular nodules of the thyroid: a morphometric study. Anal Quant Cytol Histol 13:321–328
9. Deichsel G, Trampish HJ (1985) Clusteranalyse und Diskriminanzanalyse. Fischer, Stuttgart New York
10. Doudkine A, Macaulay C, Poulin N, Palcic B (1995) Nuclear texture measurements in image cytometry. Pathologica 87: 286–299
11. Fadda G, Rabitti C, Minimo C, Ieraci A, Venzi A, Bianchi A, Lancia M, Gulotta G, Capelli A (1995) Morphologic and planimetric diagnosis of follicular thyroid lesions on the fine needle aspiration cytology. Anal Quant Cytol Histol 17:247–256
12. Ferrer-Roca O, Ballester-Guardia E, Martin-Rodriguez JA (1990) Morphometric, densitometric and flow cytometric criteria for the automated classification of thyroid lesions. Anal Quant Cytol Histol 12:48–55
13. Fleege JC, van Diest PJ, Baak JPA (1990) Computer assisted efficiency testing of different sampling methods for selective nuclear graphic Tablet morphometry. Lab Invest 63:270–275
14. Fleege JC, van Diest PJ, Baak JPA (1993) Systematic random sampling for selective interactive nuclear morphometry in breast cancer sections. Anal Quant Cytol Histol 15:281–289
15. Fonseca E, Sobrinho-Simoes M (1995) Diagnostic problems in differentiated carcinomas of the thyroid. Pathol Res Pract 191:318–341
16. Galera-Davidson H, Bibbo M, Bartels PH, Dytch HE, Gonzalez-Campora R (1986) Differential diagnosis between follicular adenoma and follicular carcinoma of the thyroid by marker features. Anal Quant Cytol Histol 8:195–200
17. Giardina C, Pollice L, Ricco R, Vacca E, Penella A, Serio G, Mastroguilio R, Potente F, Delfino VP (1989) Differential diagnosis between thyroid follicular adenoma and carcinoma: Analytic morphometric approach. Pathol Res Pract 185:726–728
18. Greenebaum E, Koss LG, Elequin F, Silver CE (1985) The diagnostic value of flow cytometric DNA measurements in follicular tumors of the thyroid gland. Cancer 56:2011–2018
19. Hedinger C (1988) Histological typing of thyroid tumors. Springer, Berlin Heidelberg New York
20. Jasani B, Schmid KW (1993) Immunocytochemistry in diagnostic histopathology. Churchill Livingstone, Edinburgh London Madrid
21. Johannessen JV, Sobrinho-Simoes M (1983) Well differentiated thyroid tumors: problems in diagnosis and understanding. Pathol Annu 18:255–284
22. Kahn NF, Perzin KH (1983) Follicular carcinoma of the thyroid: an evaluation of the histologic criteria used for diagnosis. Pathol Annu 18:221–253
23. Kriete A, Romen W, Schäffer R, Harms H, Haucke M, Gerlach B, Aus HM, ter Meulen V (1983) Computer analysis of chromatin arrangement and nuclear texture in follicular thyroid tumors. Histochemistry 78:227–230
24. Kriete A, Schäffer R, Harms H, Aus HM (1985) Computer-based cytophotometric classification of thyroid tumors in imprints. J Cancer Res Clin Oncol 56:2011–2018
25. La Rosa GL, Cavallari V, Giufridda D, Scimone S, La Porta GA, Maiorana MC, Maiorana A, Belfiore A (1990) The morphometric analysis of cell nuclei from the fine needle aspirates of thyroid follicular lesions does not improve the diagnostic accuracy of traditional cytologic examination. J Endocrinol Invest 13:701–707
26. Liautaud-Roger F, Dufer J, Pluot M, Delisle MJ, Coninx P (1989) Contribution of quantitative cytology to the cytological diagnosis of thyroid neoplasms. Anticancer Res 9:231–234
27. Liautaud-Roger F, Dufer J, Delisle MJ, Coninx P (1992) Thyroid neoplasms: can we do any better with quantitative cytology? Anal Quant Cytol Histol 14:373–378
28. Luck JB, Mumaw VR, Frable WJ (1982) Fine needle aspiration biopsy of the thyroid: differential diagnosis by videoplan image analysis. Acta Cytol 26:793–796
29. Marchevsky AM, Gil J, Jeanty H (1987) Computerized interactive morphometry in pathology: Current instrumentation and methods. Hum Pathol 18:320–331
30. Mattfeld T, Schürmann G, Feichter G (1987) Stereology and flow cytometry of well-differentiated follicular neoplasms of the thyroid gland. Virchows Arch [A] 410:433–441
31. Montironi R, Santinelli A, Scarpelli M, Braccischi A (1992) Cytologic diagnosis of non-oxyphil follicular neoplasias of the thyroid: multivariate classification procedure based on quantitative nucleolar features. Pathol Res Pract 188:504–509
32. Nafe R, Fritsch RS, Soudah B, Hamann A, Choritz H (1992) Histomorphometry in paraffin sections of thyroid tumors. Pathol Res Pract 188:1042–1048
33. Norusis MJ (1994) SPSS professional statistics 6.1. SPSS, Chicago
34. Rosai J (1996) Ackerman's surgical pathology. Mosby, St Louis Baltimore Boston
35. Rosen Y, Rosenblatt P, Saltzmann E (1990) Intraoperative pathologic diagnosis of thyroid neoplasms. Cancer 66:2001–2006
36. Salmon I, Kiss R, Franc B, Gasperin P, Heimann R, Pasteels JL, Verhest A (1992) Comparison of morphonuclear features in normal, benign and neoplastic thyroid tissue by digital cell image analysis. Anal Quant Cytol Histol 14:47–54
37. Salmon I, Gasperin P, Pasteels JL, Heimann R, Kiss R (1992) Relationship between histopathologic typing and morphonuclear assessments of 238 thyroid lesions. Am J Clin Pathol 97:776–786
38. Sassi I, Mangili F, Sironi M, Freschi M, Cantaboni A (1989) Morphometric evaluation of fine needle biopsy of single thyroid nodules. Pathol Res Pract 185:722–725
39. Schmid KW, Hofstädter F, Propst A (1986) A fourteen year practice with the fine needle aspiration biopsy of the thyroid in an endemic area. Pathol Res Pract 181:308–310
40. Schmid KW, Ladurner D, Zechmann W, Feichtinger H (1989) Clinicopathologic management of tumors of the thyroid gland in an endemic goiter area. Acta Cytol 33:27–30
41. Schröder S, Pfannschmidt N, Dralle H, Arps H, Böcker W (1984) The encapsulated follicular carcinoma of the thyroid. Virchows Arch 402:259–273
42. Schürmann G, Mattfeldt T, Feichter G, Buhr HJ (1990) Probleme der Differentialdiagnose follikulärer Neoplasien der Schilddrüse. Langenbecks Arch Chir 375:95–101
43. Schürmann G, Mattfeldt T, Feichter G, Koretz K, Möller P, Buhr H (1991) Stereology, flow cytometry, and immunohistochemistry of follicular neoplasms of the thyroid gland. Hum Pathol 22:179–184
44. Slowinska-Klencka D, Klencki M, Sporny S, Lewinski A (1993) The usefulness of computer-aided karyometric examination in preoperative differentiation of follicular neoplasms of the thyroid gland. Thyroidology 5:29–33
45. Soares P, Sobrinho-Simoes M (1995) Recent advances in cytometry, cytogenetics and molecular genetics of thyroid tumors and tumor-like lesions. Pathol Res Pract 191:304–307
46. Stanta G, Carcangiu ML, Rosai J (1988) The biochemical and immunohistochemical profile of thyroid neoplasia. Pathol Annu 23:129–151
47. Wright RG, Castles H, Mortimer RH (1987) Morphometric analysis of thyroid cell aspirates. J Clin Pathol 40:433–445