Analysis of interobserver reproducibility in grading histological patterns of dysplastic nevi*

Avaliação da reprodutibilidade interobservador da graduação de padrões histológicos dos nevos displásicos

Aretha Brito Nobre[1]
Juan Piñeiro-Maceira[2]
Ronir Raggio Luiz[3]

Abstract: BACKGROUND: Dysplastic nevi are among the most important cutaneous melanoma simulators. They are important risk markers for this neoplasia and can be its potential precursors. Some authors found a statistically significant relationship between the degree of dysplasia and the risk for developing melanoma. However, reproducibility of grading criteria ranged from poor to fair in the researched articles.
OBJECTIVE: To test the reproducibility of the grading criteria proposed by Sagebiel et al. regarding dysplastic nevi.
METHODS: Histological specimens of 75 dysplastic nevi were graded, independently and in a blinded fashion, according to pre-established criteria, by a panel of 10 pathologists with different levels of experience. Diagnostic agreement was calculated using weighted kappa and intraclass correlation coefficients.
RESULTS: The average of weighted kappa values was 0.13 for all observers, 0.12 for dermatopathologists, 0.18 for general pathologists and 0.05 for residents. Intraclass correlation coefficient values were 0.2 for all observers, 0.18 for dermatopathologists, 0.33 for general pathologists and 0.15 for residents.
CONCLUSIONS: Histopathological grading for dysplastic nevi was not reproducible in this Brazilian series, so the criteria used are not a helpful histopathological parameter for clinicopathological correlation.
Keywords: Dysplastic nevus syndrome; Pathology; Reproducibility of results

Resumo: FUNDAMENTOS: Nevos displásicos estão entre os mais importantes simuladores de melanoma. São marcadores de risco para o desenvolvimento dessa neoplasia e podem ser seus precursores. Alguns autores observaram uma relação estatisticamente significativa entre o grau de displasia e o risco de desenvolvimento de melanoma. No entanto, a reprodutibilidade dos critérios para graduação variou de ruim a razoável nos artigos consultados.
OBJETIVO: Testar a reprodutibilidade da graduação proposta por Sagebiel et al. para os nevos displásicos.
MÉTODOS: Seções histológicas de setenta e cinco nevos displásicos foram graduadas, de forma independente e anônima, segundo critérios pré-estabelecidos, por um painel de 10 patologistas com diferentes níveis de experiência. A concordância diagnóstica foi calculada usando os coeficientes de kappa ponderado e de correlação intraclasse.
RESULTADOS: A média dos valores de kappa ponderado foi de 0,13 para todos os observadores, de 0,12 para os dermatopatologistas, de 0,18 para os patologistas gerais e de 0,05 para os residentes. Os valores dos coeficientes de correlação intraclasse foram 0,2 para todos os observadores, 0,18 para os dermatopatologistas, 0,33 para os patologistas gerais e 0,15 para os residentes.
CONCLUSÕES: A graduação histopatológica dos nevos displásicos não foi reprodutível nesta série brasileira. Portanto, os critérios utilizados não são úteis para correlação clínico-patológica.
Palavras-chave: Patologia; Reprodutibilidade dos testes; Síndrome do nevo displásico

Received on 28.12.2011
Approved by the Advisory Board and accepted for publication on 25.04.2012.
* Work conducted at the Service of Anatomical Pathology - Hospital Universitário Clementino Fraga Filho (Clementino Fraga Filho University Hospital) – Federal University of Rio de Janeiro (HUCFF-UFRJ) – Rio de Janeiro (RJ), Brazil.
Financial Support: None.
Conflict of Interest: None.

[1] MD, MsC - Pathologist – Instituto Nacional de Câncer (National Cancer Institute -INCA) – Rio de Janeiro (RJ), Brazil.
[2] MD, PhD - Professor – Federal University of Rio de Janeiro - UFRJ – Rio de Janeiro (RJ), Brazil.
[3] DsC - Professor – Federal University of Rio de Janeiro - UFRJ – Rio de Janeiro (RJ), Brazil.

©2013 by Anais Brasileiros de Dermatologia

INTRODUCTION

Dysplastic melanocytic nevi are among the most important melanoma simulators, both on clinical and histopathological examination.[1-3] They are also important risk markers for melanoma and can be its potential precursors.[4, 5,6]

Given the importance of melanocytic nevi, studies have been conducted to establish criteria for grading dysplasia, as occurs with cervical intraepithelial lesions or adenomas of the gut, for example. Shors *et al*.[7] and Arumi-Uria *et al*.[8] found a statistically significant relationship between the degree of dysplasia and the risk for melanoma, and this finding could prove the usefulness of grading these lesions. However, Shors *et al*. found low agreement among the 13 dermatopathologists that participated in the study. Other authors also obtained similar results, considering the low agreement rates presented. This shows that more precise criteria and better reproducibility should be established.[9,10]

The great controversy still surrounding the histopathological diagnosis of dysplastic nevi and the possible usefulness of histological grading have sparked interest in assessing the reproducibility of grading criteria as proposed by Sagebiel *et al*. among pathologists with different levels of experience.[11] We decided to adopt the grading criteria proposed by Sagebiel *et al*. because this was the first article in our literature review in which the histopathological characteristics for this purpose were exposed in a didactic way and were accompanied by drawings.[11]

MATERIAL AND METHODS

We conducted a retrospective analysis of all histological specimens diagnosed as dysplastic nevus filed in the pathology section of Hospital Universitário Clementino Fraga Filho - UFRJ (Rio de Janeiro – Brazil), from 1978 to 2006, which resulted in a total of 94 lesions. The anatomical sites and ages of the patients at the time of removal of the lesions were recorded based on the information written in the request for histopathological examination.

Two slides corresponding to partial samples (incisional biopsies or lesions incompletely excised), two slides with artifacts with no paraffin block available and one case of acral location were excluded from the study, resulting in 89 lesions. The acral lesion was excluded because this type of lesion tends to have an atypical morphology.[2]

Histopathological evaluation

The selected cases – histological sections stained with hematoxylin and eosin – were separately evaluated by a general pathologist and a dermatopathologist, according to the criteria for diagnosis of dysplastic nevi; later, the criteria established by Sagebiel *et al*. were applied for grading melanocytic dysplasia as mild, moderate or severe.[11] The definition of cytological atypia was nuclear pleomorphism of melanocytes present in more than 90% of the extension of the nevus, based on what was observed during preliminary evaluation of the sample.

Confirmed cases of dysplastic nevi were evaluated by a panel of eight observers with different levels of experience – three dermatopathologists, three general pathologists and two third-year Pathology residents – for application of the grading criteria suggested by Sagebiel *et al*. Subsequently, we conducted an agreement analysis.[11]

Information for panel members

Each panel member received a sheet of paper with the description of the criteria and a form with the slide numbers and a provided space for "observations", where pathologists or residents could disagree about the diagnosis of dysplastic nevi. In this case, they would not grade the lesion. (Chart 1).[11]

Statistical analysis

Interobserver reproducibility was analyzed using kappa statistics. Kappa is a measure of agreement between two observers used for categorical data. The measure incorporates a correction for the extent of agreement expected by chance alone. A kappa value equal to zero indicates agreement that can be explained by chance alone, whereas a value equal to 1 indicates total agreement (100%). A negative kappa value is found when the observed agreement is less than expected by chance and shows systematic disagreement. It was suggested that a kappa value greater than 0.8 reflects almost perfect agreement, values between 0.61 and 0.8 indicate substantial agreement, values between 0.6 and 0.41 denote moderate agreement and values below 0.4 indicate low agreement. However, kappa is calculated for nominal scale, so it does not reveal how important the discrepancies are. For example, a disagreement between mild dysplasia and severe dysplasia is computed as equivalent to a disagreement between mild dysplasia and moderate dysplasia. To correct this problem, we used weighted kappa, assigning weights to the categories of the variable. Thus, we assigned weight 1 to "conventional" (non-dysplastic) melanocytic nevus or lentigo, weight 2 to melanocytic dysplastic nevi with mild dysplasia, weight 3 to dysplastic melanocytic nevus with moderate dysplasia, weight 4 to dysplastic melanocytic nevus with severe dysplasia, weight 5 to melanoma in situ and weight 6 to invasive melanoma. The intraclass correlation coef-

Chart 1: Form distributed to panel members for grading lesions

CRITERIA FOR DYSPLASIA IN MELANOCYTIC NEVI:
1. Downgrowth and complexity of the epidermal ridge pattern, with hyperplasia of melanocytes in a disordered growth.

2. Host response of perivascular lymphocytes and lamellar fibroplasia.

3. Cytologic atypia not prominent in mild and moderate categories; cytologic atypia and entrapment of dysplastic nevus cells in desmoplastic subepidermal collagen present in severe dysplasia.

GRADING OF DYSPLASIA:
Mild: Dysplastic changes focally present, usually at the periphery of the nevus.

Moderate: Dysplastic changes present over the entire extension of the nevus, but without significant cytologic atypia

Severe: Severe: Dysplastic changes present in moderate form, plus cytologic atypia (nuclear pleomorphism of melanocytes present in more than 90% of the extension of the nevus), with or without entrapment of cells within desmoplastic collagen.

Note: If there is disagreement in the diagnosis of dysplastic melanocytic nevi, the new diagnosis must be placed in the "OBS" field of the grading form and, therefore, the lesion should not be graded.

Source: Sagebiel RWet al, 1985. [11]

ficient was also calculated and it can be interpreted in the same way as weighted kappa.[1]

RESULTS

The 89 lesions were from 54 patients (27 female and 27 male) and were most often located on the back. The average age of patients at the time of lesion removal was 33.69 (7 - 86 years).

Preliminary assessment

Lesions were separately examined by a general pathologist and a dermatopathologist, according to the criteria for the diagnosis of dysplastic nevi; later, the criteria established by Sagebiel *et al*.[11] for grading melanocytic dysplasia were applied. In table 1, we can see the initial comparison between their diagnoses.

To generate a kappa value and statistical significance in SPSS 13.0, we need a table with the same number of rows and columns. Thus, some diagnoses were grouped into categories. The diagnosis of AJN/MIS and proliferation of atypical melanocytes were grouped and placed in the category "doubt". The diagnoses of non-dysplastic melanocytic nevus, Reed nevus and "no nevus" were placed in the category "conventional nevi" and the diagnoses of invasive melanoma and melanoma in situ were grouped in the category "melanoma".

Therefore, we could do an initial statistical analysis, reaching a kappa of 0.521, with p-value less than 0.001 and standard error of 0.067.

In order to use the software KAPPA.exe (PEPI) 4.0 to calculate weighted kappa, the category "doubt" was removed, resulting in 0.49, with p-value less than 0.001 and standard error of 0.066, which reflects a moderate agreement among pathologists with different levels of experience.

No progressive relationship was found between the age of the patients and the degree of dysplasia, considering the diagnoses of the dermatopathologist.

Panel results

The 75 lesions with a diagnosis of dysplastic melanocytic nevus confirmed by the dermatopathologist during the initial assessment were distributed to eight observers (Table 1). The three participating dermatopathologists were called A, B and C. The three general pathologists were called D, E and F. The two residents were called G and H. The dermatopathologist who participated in the preliminary assessment was called I and the general pathologist, J. The frequency of diagnoses for each observer is shown in table 2.

Only 72 lesions evaluated by observer D were computed because two lesions were diagnosed as "atypical lentigo" and one as "proliferation of melanocytes of lentiginous type". Since there was no category for these diagnoses, we decided to exclude them from the statistical analysis. We analyzed 73 diagnoses from resident H because both options of mild and moderate dysplasia were marked simultaneously for two lesions. We used the intraclass correlation coefficient as an estimation of kappa to calculate the overall concordance between the groups of observers, resulting in 0.20 for all observers, 0.18 for dermatopathologists, 0.33 for general pathologists and 0.15 for residents.[12]

Table 1: Diagnostic comparison between the dermatopathologist and general pathologist during preliminary assessment

		Dermatopathologist										
		CN	**MiD**	**MoD**	**SD**	**MIS**	**nonevus**	**AJN/MIS**	**RN**	**Prolifat mel**	**IM**	**Total**
General Pathologist	CN	0	2	0	1	0	0	0	0	0	0	3
	MiD	2	14	1	1	0	0	0	0	0	0	18
	MoD	2	9	32	1	0	1	0	0	0	1	46
	SD	0	5	2	6	0	0	0	0	0	1	14
	MIS	0	0	0	1	2	0	0	0	0	0	3
	AJN/MIS	0	0	0	0	0	0	1	0	0	0	1
	RN	0	0	0	0	0	0	0	2	0	0	2
	Prolifat mel	0	0	0	0	0	0	0	0	2	0	2
	Total	4	30	32	10	2	1	1	2	2	2	89

CN: "conventional" (non-dysplastic) melanocytic nevus; MiD: dysplastic melanocytic nevus with mild dysplasia; MoD: dysplastic melanocytic nevus with moderate dysplasia; SD: dysplastic melanocytic nevus with severe dysplasia; MIS: melanoma *in situ*; AJN/MIS: atypical (dysplastic) junctional nevus/ melanoma in situ; RN: Reed nevus; Prolif at mel: proliferation of atypical melanocytes ; IM: invasive melanoma.

Table 2: Frequency of diagnoses made by the different observers in the 75 cases of dysplastic nevi previously selected.

	Observers									
Diagnoses	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
CN	14 (18.7)	–	7 (9.3)	16 (21.3)	23 (30.7)	–	4 (5.3)	20 (26.7)	–	3 (4.0)
Mi	27 (36.0)	24 (32.0)	28 (37.3)	27 (36.0)	28 (37.3)	44 (58.7)	25 (33.3)	30 (40.0)	30 (40.0)	16 (21.3)
MoD	23 (30.7)	35 (46.7)	26 (34.7)	15 (20.0)	15 (20.0)	26 (34.7)	32 (42.7)	19 (25.3)	35 (46.7)	42 (56.0)
SD	10 (13.3)	16 (21.3)	6 (8.0)	14 (18.7)	9 (12.0)	5 (6.7)	14 (18.7)	4 (5.3)	10 (13.3)	13 (17.3)
MIS	1 (1.3)	–	1 (1.3)	–	–	–	–	–	–	1 (1.3)
IM	–	–	7 (9.3)	–	–	–	–	–	–	–
Total	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)

The weighted kappa values for each pair of observers are shown in table 3, as well as the values of agreement percentage, p-value and standard error. Lines 1 to 6 represent crossing between dermatopathologists, lines 7 to 21 between dermatopathologists and general pathologists, lines 22 to 27 between general pathologists, line 28 between residents, lines 29 to 36 between dermatopathologists and residents, and lines 37 to 44 between general pathologists and residents.

The average weighted kappa values for the total number of observers and also for each group were calculated, as can be seen in table 4.

As previously stated, weights were assigned to categories of the variable: category 1 – "conventional" melanocytic nevus or lentigo; category 2 - dysplastic melanocytic nevi with mild dysplasia; category 3 - dysplastic melanocytic nevus with moderate dysplasia; category 4 - dysplastic melanocytic nevus with severe dysplasia; category 5 - melanoma in situ, and category 6 - invasive melanoma. Considering the values of the categories of the variable and comparing the diagnoses of four dermatopathologists, there was a difference greater than or equal to two points for 35 lesions (46.6%). Only four lesions received the same diagnosis from these four observers, representing 5.3% of the total. The average score was obtained from the diagnosis of dermatopathologists as an attempt to reach a consensus. Values between 1.75 and 2.25 were considered as mild dysplasia, between 2.75 and 3.25 as moderate dysplasia, between 3.75 and 4.25 as severe dysplasia and equal to 4.75 as melanoma in situ. The frequency of the average scores can be seen in table 5.

Table 3: Interobserver agreement – Weighted kappa values, p-value, standard-error and percentage of agreement for each pair of observers

Pairs of observers	Weighted Kappa	P-value	Standard-error	Agreement (%)
A versus B	0.01	0.417	0.07	24.0
A versus C	0.09	0.072	0.07	29.3
B versus C	0.12	0.033	0.06	36.0
A versus I	0.09	0.104	0.07	26.6
B versus I	0.26	0.001	0.09	58.6
C versus I	0.18	0.003	0.06	41.3
A versus D	0.18	0.010	0.08	33.3
A versus E	-0.01	0.562	0.08	22.6
A versus F	0.16	0.014	0.07	42.6
A versus J	0.10	0.057	0.07	29.3
B versus D	0.14	0.018	0.07	36.1
B versus E	0.16	0.002	0.06	26.6
B versus F	0.13	0.045	0.08	44.0
B versus J	0.27	< 0.001	0.08	54.6
C versus D	0.08	0.108	0.07	30.5
C versus E	0.15	0.006	0.06	32.0
C versus F	0.09	0.070	0.06	37.3
C versus J	0.05	0.204	0.06	32.0
D versus I	0.22	0.001	0.07	37.5
E versus I	0.16	0.003	0.06	32.0
F versus I	0.14	0.051	0.09	46.6
D versus E	0.34	< 0.001	0.08	37.5
D versus F	0.12	0.048	0.07	40.2
D versus J	0.31	< 0.001	0.07	40.2
E versus F	-0.01	0.593	0.06	18.6
E versus J	0.21	< 0.001	0.06	33.3
F versus J	0.13	0.031	0.07	36.0
G versus H	0.05	0.210	0.06	27.3
A versus G	0.04	0.314	0.07	28.0
A versus H	0.22	0.002	0.08	39.7
B versus G	-0.01	0.542	0.08	33.3
B versus H	0.10	0.026	0.05	32.8
C versus G	0.11	0.053	0.07	29.3
C versus H	0.02	0.393	0.06	21.9
G versus I	0.15	0.034	0.08	40.0
H versus I	0.20	< 0.001	0.06	32.8
D versus G	-0.03	0.655	0.07	23.6
D versus H	0.35	< 0.001	0.08	48.5
E versus G	-0.09	0.910	0.06	20.0
E versus H	0.25	0.001	0.08	35.6
F versus G	0.17	0.015	0.08	44.0
F versus H	0.13	0.028	0.07	43.8
G versus J	0.12	0.055	0.08	42.6
H versus J	0.12	0.010	0.05	32.8

Eight out of 23 lesions with values equal to 2.50 and 3.50, considered as "no consensus" and corresponding to 30.7% of the total, showed a difference greater than or equal to two points. However, a consensus could be reached for five lesions because more than two dermatopathologists reached the same diagnosis. Therefore, three lesions were considered "problematic" cases.

In table 6, we present the diagnoses of each dermatopathologist for these three lesions. We can see that the same lesion received a diagnosis of dysplastic melanocytic nevi with mild dysplasia and of melanoma in situ by two different dermatopathologists, while the other two received a diagnosis of "conventional" melanocytic nevus and also dysplastic melanocytic nevus with severe dysplasia. The same strategy was used for the diagnoses of general pathologists, and the value of 1.25 was consid-

Table 4: Average values of weighted kappa

Observer groups	Average values of weighted kappa (min - max)
Dermatopathologists	0.12 (0.01 - 0.26)
Dermatopathologists x general pathologists	0.13 (-0.01 - 0.27)
General Pathologists	0.18 (-0.01 - 0.34)
Dermatopathologists x residents	0.10 (-0.01 - 0.22)
General pathologists x residents	0.12 (-0.09 - 0.35)
Total number of observers	0.13 (-0.09 - 0.35)

Table 5: Frequency of the average score for the diagnosis of dermatopathologists

Diagnosis	Average score	Number of cases	Percentage
MiD	1.15 to 2.25	20	26.6
MoD	2.75 to 3.25	27	36.1
SD	3.75 to 4.25	4	5.3
MIS	4.75	1	1.3
no consensus	2.50 and 3.50	23	30.7

MiD: dysplastic melanocytic nevus with mild dysplasia;
MoD: dysplastic melanocytic nevus with moderate dysplasia; SD: dysplastic melanocytic nevus with severe dysplasia; MIS: melanoma *in situ*.

Table 6: Cases with greater diagnostic difficulty among dermatopathologists

	Diagnosis - Dermatopathologists			
	A	**B**	**C**	**I**
Case 23	MoD	MiD	MIS	SD
Case 40	CN	SD	MoD	MiD
Case 44	CN	MiD	MoD	SD

CN: "conventional" (non-dysplastic) melanocytic nevus or lentigo; MiD: dysplastic melanocytic nevus with mild dysplasia; MoD: dysplastic melanocytic nevus with moderate dysplasia; SD: dysplastic melanocytic nevus with severe dysplasia; MIS: melanoma *in situ*

ered as "conventional" (non-dysplastic) melanocytic nevus or lentigo (Table 7). Average values were obtained from only 72 lesions because observer D diagnosed two lesions as "atypical lentigo" and one as "proliferation of melanocytes of lentiginous type", as previously mentioned.

One out of 14 lesions with values equal to 1.50, 2.50 and 3.50, considered as "no consensus" and equivalent to 19,5% of the total, showed a difference greater than or equal to two points, since it received the diagnosis of "conventional" (non-dysplastic) melanocytic nevus and dysplastic melanocytic nevus with severe dysplasia. However, there was another lesion with a difficult diagnosis which was not included in the statistics, as it was diagnosed as "atypical lentigo" by observer D, as previously stated. This one was diagnosed as "conventional" (non-dysplastic) melanocytic nevus and dysplastic melanocytic nevus with moderate dysplasia (Table 8).

Only one lesion received the same diagnosis by the four observers, representing 1.3% of the total. Twenty out of 73 lesions evaluated by the two residents received the same diagnosis and differences were greater than or equal to two points in nineteen lesions, equivalent to 27.3% and 26% of the total, respectively. No lesion was diagnosed the same way by the 10 observers.

DISCUSSION

The 89 cases initially diagnosed as dysplastic melanocytic nevus by pathologists at Hospital Universitário Clementino Fraga Filho were examined by the most experienced author in this field (observer I - a dermatopathologist) so that the sample would be as representative of dysplastic melanocytic nevi as possible. When analyzing the diagnoses made by the dermatopathologist and the general pathologist during the preliminary assessment of dysplastic melanocytic nevi, we observed a tendency of the general pathologist (obviously less experienced in melanocytic lesions than the dermatopathologist) to allocate a greater number of lesions to the moderate

Table 7: Frequency of the average score for the diagnosis of general pathologists.

Diagnosis	Average score	Number of cases	Percentage
CN	1.25	1	1.4
MiD	1.75 to 2.25	29	40.2
MoD	2.75 to 3.25	24	33.3
SD	3.75 to 4.25	4	5.6
no consensus	1.50. 2.50 and 3.50	14	19.5

CN: "conventional" (non-dysplastic) melanocytic nevus or lentigo; MiD: dysplastic melanocytic nevus with mild dysplasia; MoD: dysplastic melanocytic nevus with moderate dysplasia; SD: dysplastic melanocytic nevus with severe dysplasia; MIS: melanoma *in situ*.

category (Table 1). In the study by Duncan *et al*., the authors found that the two less experienced observers allocated a disproportionately larger number of lesions to the moderate category and indicated that this could be a "safe haven" category.[9] When comparing the initial diagnoses made by the dermatopathologist and the general pathologist, we noted that both had difficulty in the diagnosis of the same lesions (AJN/MIS and Prolif at mel), which were allocated to the category "doubt". However, there were two lesions with great diagnostic discrepancy, both diagnosed as invasive melanoma by the dermatopathologist and dysplastic melanocytic nevus by the general pathologist, one with moderate dysplasia and the other with severe dysplasia. This can also be explained by the general pathologist's lack of experience in the analysis of this type of lesion during the preliminary evaluation of the sample. Moreover, these two lesions had been diagnosed as dysplastic nevi by the hospital pathologists, and this may have influenced the general pathologist to consider them as dysplastic nevi.

In table 2, we see that dermatopathologist C was the only one to diagnose invasive melanoma (seven lesions), while dermatopathologist B diagnosed all the lesions of the sample as dysplastic nevi. Three observers (A, C and J) diagnosed melanoma in situ; however, the diagnosis was given to three different lesions. Seven observers diagnosed some lesions as "conventional" (non-dysplastic) melanocytic nevus; three observers (two general pathologists and one resident) gave this diagnosis to more than 20% of the sample. This could be explained by lack of experience in diagnosing melanocytic lesions, but dermatopathologist A also identified 18.7% of lesions as "conventional" melanocytic nevi. In addition, all observers considered the majority of the sample as dysplastic nevi, which indicates good sample representativeness. However, there may have been some bias toward dysplastic melanocytic nevus because the lesions were distributed with a confirmed diagnosis made by the more experienced dermatopathologist (observer I). As can be seen in table 3, weighted kappa values were low for almost all pairs of observers. The highest weighted kappa values were 0.35 (between a general pathologist and a resident) and 0.34 (between general pathologists), showing that the maximum reached was a fair agreement.[13] When evaluating the percentage of agreement between pairs, note that the highest value is 58.6%, a crossing between two dermatopathologists, which represents fair agreement even among experienced pathologists in this area.

Table 8: Cases with greater diagnostic difficulty among general pathologists.

	Diagnoses - General pathologists			
	D	**E**	**F**	**J**
Case 12	SD	CN	MiD	MoD
Case 63	AL	CN	MiD	MoD

CN: "conventional" (non-dysplastic) melanocytic nevus or lentigo; MiD: dysplastic melanocytic nevus with mild dysplasia; MoD: dysplastic melanocytic nevus with moderate dysplasia; SD: dysplastic melanocytic nevus with severe dysplasia; AL: atypical lentigo.

The average weighted kappa values indicate poor agreement between groups and between the total number of observers, because all were below 0.2 (Table 4). The results shown in Table 5 also confirm the good representativeness of the sample. When observing the average score for the diagnosis of dermatopathologists, we see that no lesion had a score lower than 1.75 and only one had a score equal to 4.75. Duray *et al*. commented that "all anatomic pathology diagnoses are formed by value judgments that result from the conscious interpretation of histologic imagery (pattern analysis)".[14] Therefore, there is subjectivity in the pathological analysis and, not surprisingly, there is variability in interpretation even among experienced pathologists in a particular subject, especially when borderline lesions are involved,

such as dysplastic nevi. In this study, although the same criteria were used by all observers, a great variety of diagnoses was found, resulting in low agreement. Analyzing Sagebiel´s criteria, one realizes that the only feature that differentiates severe dysplasia from the others is the presence of cytological atypia, but those authors did not define it. To reduce interobserver bias, we defined cytologic atypia as nuclear pleomorphism of melanocytes present in more than 90% of the extension of the nevus, based on what was observed during the preliminary evaluation of the sample. Other authors have used this definition with the same purpose.[15] Even so, there was low reproducibility in the identification of this feature, as reported in another study, and perhaps this is one of the reasons that may explain the low agreement.[10] However, a considerable part of diagnostic discrepancy occurred in the differentiation between mild dysplasia and moderate dysplasia. The difference between one category and the other is due to, according to Sagebiel, the presence of architectural features to the fullest extent or just a part of the nevus. Therefore, the identification of these characteristics also had low reproducibility. Intraclass correlation coefficient values were also low and can be interpreted in the same way as weighted kappa values.[12] The highest value (0.33) was the result of the diagnoses made by the general pathologists, which may be explained by greater adherence to the criteria provided. In other words, as they had less previous experience in the subject than dermatopathologists, the criteria may have been followed more strictly, as there was no comparison parameter prior to this work. Contrary to what was reported by Sagebiel *et al*., no progressive relationship was found between the age of the patients and the degree of dysplasia, although the diagnosis of only one observer – the most experienced dermatopathologist - was taken into account.[11] The results of this study are consistent with the literature regarding low reproducibility when grading dysplastic nevi. Although one study has shown fair interobserver concordance, others have shown low to moderate agreement for grading or poor reproducibility of some grading criteria.[9,10,16] Therefore, histologic grading of dysplastic nevi cannot be a determinant for clinical management, as verified by Fung in his survey of 145 dermatologists, in which the most common reason for re-excising incompletely removed dysplastic nevi was reporting of moderate (or worse) cytologic atypia by the pathologist.[17] In our view, contrary to what has been stated by Goodson *et al.*, one may not measure the rate of clinical recurrence based on the degree of dysplasia, since what one pathologist considers mild dysplasia may be considered as severe dysplasia by another and vice versa.[18] For the same reason, the degree of dysplasia in dysplastic nevi cannot be a parameter for higher or lower risk for developing melanoma.

Duncan *et al*. said that to establish whether the degree of melanocytic dysplasia has any biological relevance, dermatopathologists should consistently recognize two or more degrees of dysplasia.[9] That is, if the risk of melanoma is related to the degree of dysplasia, at least dermatopathologists have to be able to reliably recognize at least two degrees of dysplasia in melanocytic dysplastic nevi.

However, it is known that these lesions have varying degrees of dysplasia because some of them give rise to doubt (dysplastic melanocytic nevus x melanoma in situ or dysplastic melanocytic nevus x "conventional" melanocytic nevus), as already reported, while others are undoubtedly dysplastic nevi.[9] As rightly stated by Rhodes *et al*., more than twenty years ago, "the prognostic issues may not be resolved until we are able to document dysplastic nevi according to methods that permit reliable diagnostic classifications" and the problem persists nowadays.[19]

CONCLUSION

Histopathological grading for dysplastic nevi was not reproducible in this Brazilian series, so the criteria used are not a helpful histopathological parameter for clinicopathological correlation. ❑

ACKNOWLEDGMENTS:

The authors would like to thank Ana Paula Aguiar Vidal Sieiro, Paulo Marcos Nunes Valiante, Vera Lucia Nunes Pannain, Mayra Carrijo Rochael, Heliomar de Azevedo Valle, Maria de Fátima Guimarães Scotelaro Alves, Verônica Goulart Moreira and Cesar de Souza Bastos Junior.

REFERENCES

1. Blessing K. Benign atypical naevi: diagnostic difficulties and continued controversy. Histopathology. 1999;34:189-98.
2. Elder DE. Precursors to melanoma and their mimics: nevi of special sites. Mod Pathol. 2006;19:S4-S20.
3. Farrahi F, Egbert BM, Swetter SM. Histologic similarities between lentigomaligna and dysplastic nevus: importance of clinicopathologic distinction. J CutanPathol. 2005;32:405-12.
4. Rigel DS, Rivers JK, Kopf AW, Friedman RJ, Vinokur AF, Heilman ER, et al. Dysplastic Nevi. Markers for Increased Risk for Melanoma. Cancer 1989;63:86-9.
5. Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. Eur J Cancer. 2005;41:28.
6. Clark WH Jr, Reimer RR, Greene M, Ainsworth AM, Mastrangelo MJ. Origin of familial malignant melanomas from heritable melanocytic lesions: "The B-K mole syndrome". Arch Dermatol 1978;114:732.
7. Shors AR, Kim S, White E, Argenyi Z, Barnhill RL, Duray P, et al. Dysplastic naevi with moderate to severe histological dysplasia: a risk factor for melanoma. Br J Dermatol. 2006;155:988.
8. Arumi-Uria M, McNutt S, Finnerty B. Grading atypia in nevi: correlation with melanoma risk. Mod Pathol 2003;16:764.
9. Duncan LM, Berwick M, Bruijn JA, Byers HR, Mihm MC, Barnhill RL. Histopathologic recognition and grading of dysplastic melanocytic nevi: an interobserver agreement study. J Invest Dermatol 1993;100:318S-321S.
10. Hastrup N, Clemmensen OJ, Spaun E, Sondergaard K. Dysplastic nevus: histological criteria and their inter-observer reproducibility. Histopathology 1994;24:503-9.
11. Sagebiel RW, Banda PW, Schneider JS, Crutcher WA. Age distribution and histologic patterns of dysplastic nevi.J AmAcadDermatol.1985;13:975-82.
12. Luiz RR. Métodos estatísticos em estudos de concordância. In: Medronho RA, Bloch KV, Luiz RR, Werneck GL, editors. Epidemiologia.2nd ed. São Paulo: Atheneu; 2008. p.343.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
14. Duray PH, DerSimonian R, Barnhill R, Stenn K, Ernstoff MS, Fine J, et al. An analysis of interobserver recognition of the histopathologic features of dysplastic nevi from a mixed group of nevomelanocytic lesions. J Am AcadDermatol 1992;27:741-9.
15. Anessi G, Cattaruzza MS, Abeni D, Baliva G, Laurenza M, Mancchini V et al. Correlation between clinical atipia and histologic dysplasia in acquired melanocytic nevi. J Am AcadDermatol. 2001;45:77-85.
16. Piepkorn MW, Barnhill RL, Cannon-Albright LA, Elder DE, Goldgar DE, Lewis CM et, al. A multiobserver, population-based analysis of histologic dysplasia in melanocytic nevi. J Am AcadDermatol 1994;30:707-14.
17. Fung MA. Terminology and management of dysplastic nevi: responses from 145 dermatologists. Arch Dermatol 2003;139:1374-5.
18. Goodson AG, Florell SR, Boucher KM, Grossman D. Low rates of clinical recurrence after biopsy of benign to moderately dysplastic melanocytic nevi. J Am AcadDermatol 2010,62:591-6.
19. Rhodes AR, MihmJr, MC, Weinstock MA. Dysplastic melanocytic nevi: a reproducible histologic definiation emphasizing cellular morphology. Mod Pathol 1989;2:306-19.

Endereço para correspondência / Mailing address:
Aretha Brito Nobre
Serviço de Anatomia Patológica
Av. Prof. Rodolpho Paulo Rocco, 255/ Sub-solo
Ilha do Fundão
21941 913 Rio de Janeiro RJ
E-mail: arethanobre@yahoo.com.br

How to cite this article: Nobre AB, Piñeiro-Maceira J, Luiz RR. Analysis of interobserver reproducibility in grading histological patterns of dysplastic nevi. An Bras Dermatol. 2013;88(1):23-31.