PERGAMON

European Journal of Cancer 39 (2003) 1861–1865

European Journal of Cancer

www.ejconline.com

Interobserver reproducibility of ulceration assessment in primary cutaneous melanomas

A. Spatz[a,*], M.G. Cook[b], D.E. Elder[c], M. Piepkorn[d], D.J. Ruiter[e], R.L. Barnhill[f] on behalf of the EORTC Melanoma group

[a]*Institut Gustave-Roussy, Villejuif, France*
[b]*Royal Surrey County Hospital, Guildford, UK*
[c]*University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[d]*University of Washington, Seattle, WA, USA*
[e]*University Medical Center St-Radboud, Nijmegen, The Netherlands*
[f]*The George Washington Medical Center, Washington, DC, USA*

Received 29 January 2003; accepted 21 February 2003

Abstract

In the recently revised melanoma staging system proposed by the American Joint Committee on Cancer (AJCC), ulceration assessment by the pathologist is a pivotal parameter. Patients upstaged because of ulceration might be included in adjuvant trials conducted in AJCC stage II melanoma patients. Therefore, accuracy based on interobserver reproducibility for melanoma ulceration assessment is crucial for proper clinical management. In some cases, it is extremely difficult, even for an experienced pathologist, to distinguish between trauma-induced ulceration, artifact and tumoral ulceration. Whether this difficulty may be resolved by the use of a more precise definition of ulceration has not been evaluated. Therefore, we have proposed a refined definition of melanoma ulceration and we tested whether this definition might improve the interobserver interpretative reproducibility of ulceration in primary cutaneous melanomas. The results of this study support the need for a more precise definition of melanoma ulceration that rules out biopsy trauma or processing artifact and could be incorporated into a standardised pathology worksheet for reporting primary melanomas.

Keywords: Melanoma; Ulceration; Pathology; Reproducibility; Kappa statistics

1. Introduction

Ulceration of a cutaneous melanoma was first identified as an adverse prognostic feature in 1953 by Allen and Spitz in their classic description of melanoma histopathology [1]. They suggested that spontaneous acantholysis of the epidermis might reflect intrinsic properties of the melanoma invading the dermis. This prognostic significance of melanoma ulceration was corroborated by Tompkins that same year [2] and by many others since that time [3–8]. In the recent melanoma staging system proposed by the American Joint Committee on Cancer (AJCC), ulceration assessment by the pathologist is a pivotal parameter and is included in all categories of tumour evaluation [9]. The introduction by the AJCC of ulceration as a key factor for microstaging in melanomas is based on its reported independent prognostic value [10]. Melanoma ulceration upstages the patient by one subcategory (IA to IB, IIA to IIB) or even may upstage the patient from stage I to stage II for localised melanomas measuring 1.01–2.00 mm in thickness (T2) [9]. Approximately 25% of these patients may be upstaged because of ulceration [10] and might then be included in adjuvant trials conducted in AJCC stage II melanoma patients. Therefore, accurate diagnosis based on a precise definition and excellent interobserver reproducibility for melanoma ulceration assessment is crucial for the proper clinical management of melanoma patients.

* Corresponding author. Tel.: +33-1-4211-4462; fax: +33-1-4211-5263.
E-mail addresses: spatz@igr.fr (A. Spatz).

0959-8049/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0959-8049(03)00325-3

The interpretation of melanoma ulceration is reported to be one of the most reproducible of all the major histopathological features [11,12]. However, interobserver agreements for the presence of ulceration reported in literature vary with kappa values ranging from 0.56 to 0.87 [11–13]. For instance, in one study, with a reported 'good' kappa score, the interobserver proportion of disagreement was 18% meaning that a substantial category of patients would have been mis-staged in the new AJCC staging system [12]. These contrasting results in interobserver reproducibility for melanoma ulceration assessment may be due to the lack of an evidence-based definition of melanoma ulceration. In the AJCC staging system, melanoma ulceration is defined as "an absence of an intact epidermis overlying a major portion of the primary melanoma based on microscopic examination of the epidermis" [9]. In some cases, however, this definition does not permit even an experienced pathologist to distinguish between trauma-induced "absence of an intact epidermis" and tumoral ulceration [14,15]. Whether this difficulty may be better addressed by the use of a more restrictive evidence-based definition of ulceration needs to be evaluated. Therefore, in this study we have proposed a precise definition of melanoma ulceration and we have tested whether this definition might improve the interobserver reproducibility of ulceration assessment in primary cutaneous melanomas among six pathologists experienced in diagnosing melanocytic lesions.

2. Patients and methods

2.1. Materials

The study group materials were derived from 200 consecutive excisional biopsies or resections of primary cutaneous melanomas, excluding melanomas *in situ*, from the Institut Gustave-Roussy. Only paraffin sections from tissue fixed in neutral-buffered formalin were evaluated. Patient confidentiality was maintained throughout the study by using only codes. For each case, all haematoxylin- and eosin-stained slides were reviewed by a pathologist who did not participate in the study to select the slide with the thickest part of the tumour measured according to Breslow.

A first study set of 100 slides from 100 cases was randomly selected from the study group materials. The Breslow indices of the lesions on the slides ranged from 0.3 to 6.0 mm (median: 1.2 mm). These slides were assembled and distributed to six pathologists highly experienced in looking at melanocytic lesions. In this first round, each observer assessed the presence or absence of ulceration according to their own experience and interpretation of the published literature. The pathologists were invited to add comments on each specimen if desired. The observers examined the first set consecutively and independently, without knowledge of the other pathologists' assessment of ulceration.

A second study set constituted the remaining 100 slides that had not been previously evaluated by the observers. Breslow indices ranged from 0.4 to 5.2 mm (median: 1.3 mm). The 200 slides from both study sets were mixed randomly and labelled consecutively. These slides were evaluated by the same observers 3 months after the first evaluation. In this second round, each observer had to assess the presence or absence of ulceration according to the following definition: 'melanoma ulceration is defined as the combination of the following features: full-thickness epidermal defect (including absence of *stratum corneum* and basement membrane), evidence of host response (i.e. fibrin deposition, neutrophils), and thinning, effacement or reactive hyperplasia of the surrounding epidermis'. The observers examined the second set consecutively and independently, without either knowledge of the other pathologists' assessment of ulceration or the results of the evaluation of the first set.

2.2. Statistical analysis

For the statistical analysis, the Statistical Package for the Social Sciences (SPSS) for Windows (version 11.0.1 (2001), SPSS, Inc., Chicago, IL, USA) was used. Kappa statistics were used for assessment of interobserver agreement [16]. This method includes a correction for chance agreement. Interobserver reproducibility has been inferred in the published literature to be 'very good' to 'excellent' when the kappa statistics exceed 0.80 [17–19].

3. Results

The scoring of ulceration as a percentage of all slides reviewed by each pathologist varied from 13 to 29% (average: 20%) in round 1, and from 16 to 24% (average: 18%) in round 2. There was congruence of all pathologists on the presence or absence of ulceration for 72% of the cases in round 1 and 89% of the cases in round 2. Conversely, at least two pathologists disagreed with the others on 18% of the cases in round 1 and 7% of the cases in round 2. For 6 cases, spontaneous comments from the observers emphasised the difficulty of differentiating between tumoral ulceration and artifacts or scratching (Fig. 1).

Chance-adjusted agreement (kappa statistics) for the ulceration assessment during the first round and the crude percentages of melanomas considered as ulcerated per observer are summarised in Table 1. Kappa statistics for the ulceration assessment during the second round and the number of melanomas considered as

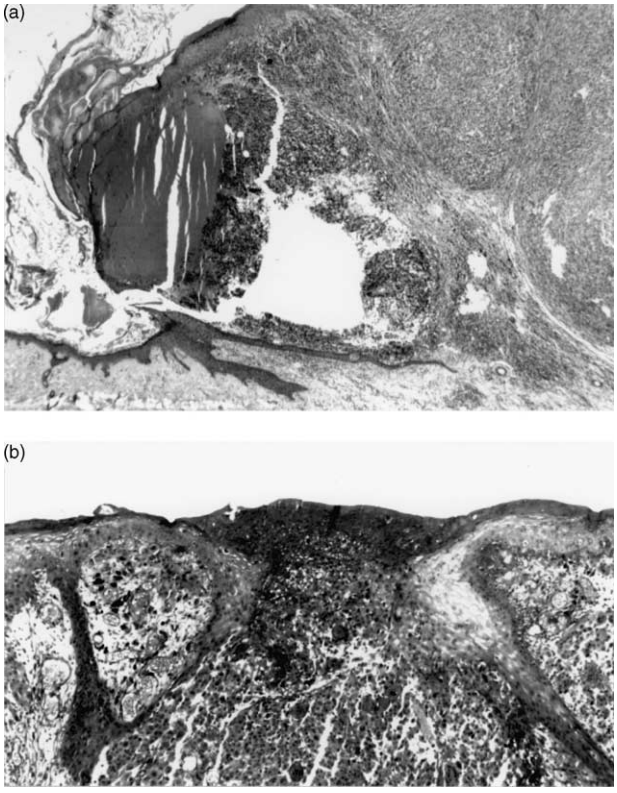

Fig. 1. (a) Study case illustrating the difficulty of assessing correctly tumoral ulceration in some cases. During the first round, all the observers spontaneously emphasised the difficulty of differentiating between tumoral ulceration and traumatic ulceration in this case that was therefore associated with a poor interobserver reproducibility for ulceration assessment. However, the lack of modifications of the surrounding epidermis argues in favour of the latter. In the second round, this case was unanimously considered as non-ulcerated. (b) By contrast, obvious ulceration is present here defined as the combination of a full-thickness epidermal defect, evidence of a host response, and modifications of the surrounding epidermis.

ulcerated per observer are provided in Table 2. In round 1, kappa scores varied from 0.47 to 0.79. The average kappa among observers was 0.63 in the first round and 0.80 in the second round. When the cases included in both rounds were considered, the average kappa among observers was 0.74. When the 'new' cases only (not included in round 1) were analysed, the average kappa was 0.86.

4. Discussion

Reproducibility of parameter assessment is but one aspect of tumour staging. Because melanoma ulceration has been shown in many studies to be a major and independent prognostic parameter, it has became a pivotal variable for the clinical management of patients with melanoma and inclusion criteria in clinical trials for adjuvant therapy [10,20–28]. Therefore, it is crucial to assure the best interobserver interpretation of this

Table 1
Interobserver interpretative reproducibility kappa statistics correcting for first round (presence of ulceration); each observer is compared with the others

Observer number	2	3	4	5	6	Number of ulcerated cases
1	0.70	0.47	0.74	0.52	0.79	29
2		0.60	0.73	0.66	0.73	20
3			0.53	0.63	0.53	13
4				0.52	0.77	23
5					0.58	16
6						23

Average among observers: 0.63.

Table 2
Interobserver interpretative reproducibility kappa statistics correcting for second round (presence of ulceration); each observer is compared with the others

Observer number	2	3	4	5	6	Number of ulcerated cases (all cases/cases included in round 1 only)
1	0.80	0.78	0.87	0.71	0.83	42/23
2		0.87	0.73	0.83	0.92	32/14
3			0.72	0.87	0.87	33/15
4				0.69	0.77	49/26
5					0.84	26/13
6						34/17

Average among observers: 0.80.

variable to avoid inappropriate treatments based on traumatic ulceration or inclusion bias in clinical trials. This is especially true for localised melanomas measuring 1.01–2.0 mm in thickness when ulceration may upstage the patient from stage I to stage II [9].

Very little peer-reviewed literature has been published addressing interobserver reproducibility for ulceration assessment in primary cutaneous melanoma and no evidence-based definition for melanoma ulceration is available. This contrasts with the importance of melanoma ulceration as a significant parameter for melanoma staging and inclusion of the patients in therapeutic protocols. Corona and colleagues reported results for 84 cases evaluated by four observers [11]. An overall agreement of 0.88 and a chance-corrected interobserver agreement (kappa) of 0.91 were reported for ulceration. In contrast, Lock-Andersen and colleagues reported an interobserver agreement of 0.82 and a kappa score of 0.65 for ulceration, compared with a chance-corrected agreement of 0.81 for thickness [12]. However, neither of these studies provided a specific definition for melanoma ulceration.

In the first round of the present study, the rate of agreement among observers was 72%, with an average kappa score of 0.63, ranging from 0.47 to 0.79. This

broad range of kappa scores was not due to a single-observer effect as at least two observers disagreed for 18% of the cases. In this first round, melanoma ulceration was defined according to the literature which was proposed in the final AJCC staging system for melanoma. as ''the absence of an intact epidermis overlying a major portion of the primary melanoma based on microscopic examination of the histological sections''. This definition does not discriminate between ulceration due to intrinsic biological properties of the tumour and extrinsic or artifactual causes for ulceration such as biopsy trauma, scratching, or technical artifacts during the gross dissection or the preparation of the slides. As a second step, we used a more refined definition (see Patients and methods). Using this definition, the rate of the cases in agreement rose to 89% and the median kappa score from 0.63 to 0.80. At the same time, the proportion of the cases interpreted differently by at least two observers decreased from 18 to 7%. The average number of ulcerated cases in each round was not significantly changed by the definition used in each round. Therefore, the definition used for the second round improved the interobserver agreement for melanoma ulceration assessment without changing the average number of ulcerated cases in the population. Such a definition might be incorporated into a standardised pathology worksheet for reporting primary melanoma.

In a retrospective analysis of 135 histopathologic reports of primary cutaneous melanomas, only 28% of them contained a statement about the presence or absence of ulceration [29]. In 50% of the cases in which ulceration was found after review, the corresponding initial pathology report had failed to mention it. The latter study suggests the need to better educate pathologists about reporting on melanomas. The definition proposed here represents an attempt to refine the diagnosis of ulceration by ruling out examples of epithelial discontinuity that are artifacts of the biopsy specimen preparation process. Finally, it must be acknowledged that traumatic 'ulceration' and ulceration related to the intrinsic properties of the tumour undoubtedly have some interrelationship, at least in some proportion of cases. The possible relationship between trauma and ulceration cannot be dismissed until the biological basis of tumour ulceration is better understood.

References

1. Allen AC, Spitz S. Malignant melanoma. A clinico-pathological analysis of the criteria for diagnosis and prognosis. *Cancer* 1953, **6**, 1–45.
2. Thompkins VN. Cutaneous melanoma: ulceration as a prognostic sign. *Cancer* 1953, **6**, 1215–1218.
3. Balch CM, Wilkerson JA, Murad TM, Soong SJ, Ingalls AL, Maddox WA. The prognostic significance of ulceration of cutaneous melanoma. *Cancer* 1980, **45**, 3012–3017.
4. Day Jr. CL, Lew RA, Harrist TJ. Malignant melanoma prognostic factors 4: ulceration width. *J Dermatol Surg Oncol* 1984, **10**, 23–24.
5. Hanekom GS, Stubbings HM, Johnson CA, Kidson SH. The detection of circulating melanoma cells correlates with tumour thickness and ulceration but is not predictive of metastasis for patients with primary melanoma. *Melanoma Res* 1999, **9**, 465–473.
6. Larsen TE, Grude TH. A retrospective histological study of 669 cases of primary cutaneous malignant melanoma in clinical stage I. 4. The relation of cross-sectional profile, level of invasion, ulceration and vascular invasion to tumour type and prognosis. *Acta Pathol Microbiol Scand [A]* 1979, **87A**, 131–138.
7. McGovern VJ, Shaw HM, Milton GW, McCarthy WH. Ulceration and prognosis in cutaneous malignant melanoma. *Histopathology* 1982, **6**, 399–407.
8. Weidner F, Schroll S, Schonberger A. The influence of lymph node metastasis and ulceration of primary melanoma on germinal centers within draining lymph nodes: a histomorphometric study. *Arch Dermatol Res* 1982, **272**, 155–161.
9. Balch CM, Buzaid AC, Soong SJ, *et al*. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol* 2001, **19**, 3635–3648.
10. Balch CM, Soong SJ, Gershenwald JE, *et al*. Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. *J Clin Oncol* 2001, **19**, 3622–3634.
11. Corona R, Mele A, Amini M, *et al*. Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. *J Clin Oncol* 1996, **14**, 1218–1223.
12. Lock-Andersen J, Hou-Jensen K, Hansen JP, Jensen NK, Sogaard H, Andersen PK. Observer variation in histological classification of cutaneous malignant melanoma. *Scand J Plast Reconstr Surg Hand Surg* 1995, **29**, 141–148.
13. Heenan PJ, Matz LR, Blackwell JB, *et al*. Inter-observer variation between pathologists in the classification of cutaneous malignant melanoma in western Australia. *Histopathology* 1984, **8**, 717–729.
14. Ruiter DJ, Spatz A, van den Oord JJ, Cook MG. Pathologic staging of melanoma. *Semin Oncol* 2002, **29**, 370–381.
15. Ruiter DJ, Spatz A, van den Oord JJ, Cook MG. Microstaging in cutaneous melanoma. *J Pathol* 2001, **195**, 525–529.
16. Cohen A. Comparison of correlated correlations. *Stat Med* 1989, **8**, 1485–1495.
17. Hoang MP, Sahin AA, Ordonez NG, Sneige N. HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. *Am J Clin Pathol* 2000, **113**, 852–859.
18. Thompson JR. Estimating equations for kappa statistics. *Stat Med* 2001, **20**, 2895–2906.
19. Tubbs R, Skacel M, Pettay J, *et al*. Interobserver interpretative reproducibility of GOLDFISH, a first generation gold-facilitated autometallographic bright field in situ hybridization assay for HER-2/neu amplification in invasive mammary carcinoma. *Am J Surg Pathol* 2002, **26**, 908–913.
20. Eggermont AM, Keilholz U, Autier P, Ruiter DJ, Lehmann F, Lienard D. The EORTC Melanoma Group: a comprehensive melanoma research programme by clinicians and scientists. European Organization for Research and Treatment of Cancer. *Eur J Cancer* 2002, **38**(Suppl. 4), S114–S119.
21. Cochran AJ, Elashoff D, Morton DL, Elashoff R. Individualized prognosis for melanoma patients. *Hum Pathol* 2000, **31**, 327–331.
22. Lang PG. Current concepts in the management of patients with melanoma. *Am J Clin Dermatol* 2002, **3**, 401–426.
23. Bartoli C, Bono A. Comments on Management of cutaneous melanoma M0: state of the art and trends, Rossi et al. Eur J Cancer, 1997, 33, 2302–2312. *Eur J Cancer* 1998, **34**, 1467–1468.
24. Eggermont AM. Strategy of the EORTC-MCG trial programme

for adjuvant treatment of moderate-risk and high-risk melanoma. *Eur J Cancer* 1998, **34**(Suppl. 3), S22–S26.
25. Eggermont AM. The current EORTC Melanoma Cooperative Group adjuvant trial programme on malignant melanoma: prognosis versus efficacy, toxicity and costs. *Melanoma Res* 1997, **7**(Suppl. 2), S127–S131.
26. Buzaid AC, Anderson CM. The changing prognosis of melanoma. *Curr Oncol Rep* 2000, **2**, 322–328.
27. Cascinelli N, Marubini E, Morabito A, Bufalino R. Prognostic factors for stage I melanoma of the skin: a review. *Stat Med* 1985, **4**, 265–278.
28. Cherpelis BS, Haddad F, Messina J, *et al*. Sentinel lymph node micrometastasis and other histologic factors that predict outcome in patients with thicker melanomas. *J Am Acad Dermatol* 2001, **44**, 762–766.
29. Busam KJ. Lack of relevant information for tumor staging in pathology reports of primary cutaneous melanoma. *Am J Clin Pathol* 2001, **115**, 743–746.