

Inherent Grading Characteristics of Individual Pathologists Contribute to Clinically and Prognostically Relevant Interobserver Discordance Concerning Broders' Grading of Penile Squamous Cell Carcinomas

Sven Gunia^a Maximilian Burger^b Oliver W. Hakenberg^c Dieter May^d
Stefan Koch^e Anjun Jain^e Kristina Birnkammer^f Wolf F. Wieland^b
Wolfgang Otto^b Ferdinand Hofstädter^g Hans-Martin Fritsche^b Stefan Denzinger^b
Christian Gilfrich^f Sabine Brookman-May^b Matthias May^f

^aInstitute of Pathology, Johanniter Hospital Stendal, Stendal, ^bDepartment of Urology, University of Regensburg, Caritas St. Josef Medical Center, Regensburg, ^cDepartment of Urology, University of Rostock, Rostock, ^dInstitute of Pathology, DRK Clinic Berlin-Köpenick, Berlin, ^eInstitute of Pathology, HELIOS Clinic Bad Saarow, Bad Saarow, ^fDepartment of Urology, St. Elisabeth Clinic Straubing, Straubing, and ^gInstitute of Pathology, University of Regensburg, Regensburg, Germany

Key Words

Penile squamous cell carcinoma · Broders' grading system · Interobserver reproducibility · Prognostic evaluation

Abstract

Introduction: We assessed the reproducibility and prognostic impact of the Broders' grading system (BGS) in a cohort of 147 patients with surgically treated penile squamous cell carcinomas. **Materials and Methods:** Conventionally stained histology slides were graded according to the BGS in two rounds by two study pathologists. Reproducibility was assessed using κ statistics. Multivariable analyses were calculated to predict cancer-specific survival (CSS). The 'mean grade' per pathologist per round was calculated by allocating grade points to each study case (G1–G4: 1–4 points) and dividing the sum of all grade points by the number of cases examined. **Results:** The BGS showed substantial interobserver variation (59–87% with $\kappa = 0.38$ – 0.69) but almost perfect intraobserver reproducibility (91% with $\kappa = 0.86$ and 96% with $\kappa = 0.94$, respectively). The 'mean grade' per pa-

thologist remained nearly constant in both rounds of examination (differences ≤ 0.05 grade points) but differed between the two pathologists (up to 0.4 grade points). In multivariable analyses, the prognostic impact of the BGS in terms of CSS was strongly pathologist-dependent. **Conclusions:** Clinically and prognostically relevant interobserver discordance concerning the BGS seems, at least in part, to be attributable to inherent 'aggressive' versus 'reserved' grading characteristics of individual pathologists.

Copyright © 2012 S. Karger AG, Basel

Introduction

In surgical pathology practice, reproducibility of the histologic grading of penile squamous cell carcinomas (PSCCs) might be substantially confounded by the well-known problem of interobserver variation [1].

However, the clinical importance conveyed by the histologic grading of squamous cell carcinomas rendered by the pathologist cannot be overstressed as it not only rep-

Table 1. Pertinent clinical and pathological characteristics of the study cohort (n = 147)

Characteristic	
Age at surgery ¹ , years	66.5 ± 11.8
Treatment of primary tumor	
Total penectomy	35 (23.8)
Partial penectomy	112 (76.2)
Tumor size ² , cm	1.9 (1.2–2.6)
Tumor thickness	
≤5 mm	48 (32.7)
>5 mm	99 (67.3)
LVI present (L1 and/or V1)	54 (36.7)
2010 pathological primary tumor classification ³	
pTa/pTis	5 (3.4)
pT1	74 (50.3)
pT2	38 (25.9)
pT3	30 (20.4)
pT4	0
2010 pathological regional lymph node involvement	
pN0/pNx	119 (80.9)
pN1	13 (8.9)
pN2	15 (10.2)
Tumor grade (local pathology)	
TG1	44 (29.9)
TG2	89 (60.5)
TG3	14 (9.5)
TG4	0

Values represent numbers of patients and percentages in parentheses, except where indicated otherwise.

¹ Mean ± SD. ² Median and interquartile range. ³ Consensus decision.

resents an independent prognostic parameter by itself, but also might strongly influence clinical management. For example, a patient with pT1aG2 PSCC will usually be spared inguinal lymph node dissection (ILND) unless the inguinal nodes appear to be involved clinically, whereas assigning the G3 category to that case would lead to upstaging to pT1b disease, which necessitates ILND, a procedure not to be underestimated in terms of possible comorbidity [2–9].

To date, only one study has assessed the reproducibility of histologic grading in 75 patients with surgically treated PSCCs, but unfortunately, it used different grading schemes which are not directly comparable [1].

We present the first comprehensive multicenter study looking at the interobserver reproducibility of the Broders' grading system (BGS) in a study cohort enrolling 147 patients with surgically treated PSCC. Moreover, by calculating the 'mean grade' per pathologist and using

multivariable Cox models for predicting cancer-specific survival (CSS), we sought to determine the magnitude of the prognostic impact conveyed by the interobserver discordance with respect to the BGS.

Methods

Study Patients

Our previously evaluated study cohort was expanded by another 55 study cases retrieved from the files of the Department of Pathology, University Hospital Regensburg, Germany [10]. Hence, a total of 147 consecutive patients surgically treated for PSCC between 1993 and 2010 in 7 different German hospitals (35 total and 112 partial penectomies) were evaluated in the present study. At surgery, all patients were free of distant metastasis (cM0) and had histologically uninvolved surgical margins. Of the 147 patients, 48 (33%) underwent ILND on the basis of clinically enlarged inguinal nodes and/or tumor stage and grade. By these indications standardized across the hospitals involved, 99 patients (67%) were spared ILND. Mean patient age at surgery was 66.5 years (range 29–92, SD 11.8) (table 1).

Pathological processing of penectomies followed similar anatomical landmarks [11]. Tumor thickness (vertical depth of infiltration from the granular cell layer of the adjacent epithelium or from the base of the ulcer to the deepest point of invasion) and lymphovascular invasion (LVI), defined as the unequivocal conventional histologic presence of tumor emboli within endothelium-lined spaces, were recorded. Patients were dichotomized based on tumor thickness (cutoff 5 mm), as previously reported elsewhere [12]. Grading was performed according to the BGS since the latter was consistently employed by the 7 pathology departments involved in this study [13]. Staging and follow-up were performed according to the guidelines of the European Association of Urology [8]. After institutional review board approval was obtained from the Medical Ethics Committee of the federal state of Brandenburg, Germany (Landesärztekammer Brandenburg; MEC No. 37588/11), CSS was obtained for all study patients.

Central Histopathologic Review and Reexamination of the Study Cases

All retrieved slides obtained from 147 study specimens (all histologically confirmed primary PSCCs) were reexamined by one study pathologist (S.G.), who selected representative study slides (median 3, range 1–6 slides) from each study case to be forwarded to a second independent study pathologist (D.M.) and reevaluated stage, LVI and tumor thickness while blinded to the retrieved original pathology reports. The second study pathologist reviewed the selected study slides for stage, LVI and tumor thickness. The two study pathologists underwent their training in different institutions and never practiced together. No subdivision of pT1 stage was performed to avoid potential confounding bias. Four study cases were assigned discrepant stages by the study pathologists. These cases were examined by a third clinical pathologist (S.K.), and the consensus pT stage assigned by the majority of raters was recorded. Moreover, based on a consensus decision, one additional patient had LVI. Regarding histologic evaluation of all other parameters, there was complete concordance amongst both study pathologists.

Then, both study pathologists held a joint meeting in which they examined 18 randomly selected PSCC specimens using a multihead microscope and discussed the Broders' criteria to be employed. A few weeks later, all study slides were circulated in two rounds to both study pathologists for histologic grading based upon the BGS. Both raters were blinded to patient ID, clinical data, stage and grade rendered in the original pathology reports, as well as the diagnostic categories previously allocated by themselves and by the other study pathologist.

Statistical Analysis

The Shapiro-Wilk normality test was used to investigate normal distribution of continuous variables. Continuous, evenly distributed variables are reported as means \pm SD. Continuous non-normally distributed variables are presented as medians with interquartile ranges. For normally distributed variables, Student's *t* test was applied. The Mann-Whitney test was used for non-normally distributed variables. The individual 'mean grade' of both study pathologists was calculated separately for their first and second round of histologic examination by allocating points to the different grading categories (G1–G4: 1–4 points) and by dividing the sum of all grade points by 147 as previously reported elsewhere [14]. In contrast, concerning the 7 independent local pathologists, their individual 'mean grade' was calculated based upon the sum of the grading categories rendered in their original pathology reports divided by the number of study cases retrieved from their corresponding files [14].

Cohen's weighted κ statistics were employed to assess reproducibility. CSS obtained from death certificates was defined as the end point to be incorporated in uni- and multivariable analyses and was calculated by the Kaplan-Meier method as the period of time between primary surgery and cancer-related death. The log-rank test was used to compare the survival curves. Univariable analyses were performed in order to assess the prognostic impact of every recorded parameter with respect to CSS. Moreover, five different multivariable Cox proportional hazards regression models were created in order to statistically evaluate the prognostic impact conveyed by the histologic grades rendered by the local pathologists as well as by the raters A and B during their first and second round of histologic examination with respect to CSS.

The build-up quality of the different multivariable Cox models was assessed by using the omnibus test in order to compare the model coefficients. Briefly, compared with a basic model (Cox model with exclusion of histologic grade), the changing deviance (-2 multiplied log-likelihood) was analyzed by means of the development of χ^2 (percentage change of model quality including the affiliated *p* values compared with the multivariable Cox model with incorporation of histologic grade). The contribution of the models' quality accuracy conveyed by those variables retained in the different models was analyzed by means of the 'area under curve' (AUC), substituted by Harrell's concordance index. The significance of the differences in accuracy estimates was tested using the Mantel-Haenszel test, which compared related AUCs [15, 16].

Statistical analyses were performed using SPSS® 17.0 (SPSS Inc., Chicago, Ill., USA) and StatView®, version 5.0 (SAS Institute Inc., Cary, N.C., USA). *p* values are two-sided, with the statistical significance level set at *p* < 0.05.

Table 2. Comparison of the Broders' grading categories allocated to the study cases (*n* = 147) by both study pathologists (raters A and B) in both rounds of examination

	TG1 (rater B)	TG2 (rater B)	TG3 (rater B)	Total
First round				
TG1 (rater A)	44	0	0	44
TG2 (rater A)	31	34	0	65
TG3 (rater A)	0	22	16	38
Total	75	56	16	147
Second round				
TG1 (rater A)	36	0	0	36
TG2 (rater A)	38	36	0	74
TG3 (rater A)	0	21	16	37
Total	74	57	16	147

TG1–3 = Tumor grades G1–G3 (Broders' grade G4 did not occur in the present series).

Results

Interobserver Reproducibility of Histopathologic Parameters

Comparison of the grading categories assigned by the 7 local pathologists and by both study pathologists showed discrepancies between G1/2 and G2/3 only (table 2). Conformity varied between 59.2 and 87.7%. κ values ranged between 0.378 and 0.692 (table 3). Regarding classification of tumor stage, LVI and dichotomized tumor thickness, both study pathologists showed complete concordance.

Intraobserver Reproducibility and 'Mean Grade' per Pathologist

Both study pathologists had almost perfect intraobserver agreement between their first and second round (91.1% and κ = 0.862 vs. 96.5% and κ = 0.942; table 3), which resulted in a nearly constant individual 'mean grade' in both rounds (rater A: 1.96 vs. 2.01, *p* = 0.575; rater B: 1.60 vs. 1.61, *p* = 0.932; table 4). However, the 'mean grades' assigned by the 7 independent local pathologists and by both study pathologists differed significantly, as did the 'mean grades' assigned by both study pathologists (*p* values < 0.001; table 4; fig. 1, 2). Whereas study pathologist A (highest 'mean grade') and B (lowest 'mean grade') showed the poorest interobserver agreement, the highest interobserver concordance was found when the study pathologists A and B were compared with the 7 local pathologists. These data indicate that the mag-

Table 3. Agreement between both study pathologists (raters A and B) and 7 independent local pathologists with respect to histologic grading

	Local pathologists	Rater A		Rater B	
		1st round	2nd round	1st round	2nd round
Local pathologists	–	87.7%, $\kappa = 0.692$	73.4%, $\kappa = 0.556$	69.4%, $\kappa = 0.495$	68.7%, $\kappa = 0.482$
Rater A (1st round)	87.7%, $\kappa = 0.692$	–	91.1%, $\kappa = 0.862$	63.9%, $\kappa = 0.446$	63.2%, $\kappa = 0.435$
Rater A (2nd round)	73.4%, $\kappa = 0.556$	91.1%, $\kappa = 0.862$	–	59.2%, $\kappa = 0.378$	59.9%, $\kappa = 0.386$
Rater B (1st round)	69.4%, $\kappa = 0.495$	63.9%, $\kappa = 0.446$	59.2%, $\kappa = 0.378$	–	96.5%, $\kappa = 0.942$
Rater B (2nd round)	68.7%, $\kappa = 0.482$	63.2%, $\kappa = 0.435$	59.9%, $\kappa = 0.386$	96.5%, $\kappa = 0.942$	–

Results are presented separately for the first and second grading round performed by both study pathologists as conformity (percentage) and κ value.

Table 4. Prognostic implications conveyed by the ‘mean grade’ in terms of CSS after 3 years

	Mean grade \pm SE	TG1, %	TG2, %	TG3, %	χ^2 , p value (log-rank)
LPs	1.80 \pm 0.05	94	62	31	17.0, $p < 0.001$
Rater A/1	1.96 \pm 0.06	98	60	52	17.9, $p < 0.001$
Rater A/2	2.01 \pm 0.06	100	64	50	17.5, $p < 0.001$
Rater B/1	1.60 \pm 0.05	89	63	18	35.6, $p < 0.001$
Rater B/2	1.61 \pm 0.06	88	64	18	35.3, $p < 0.001$

LPs = Local pathologists. See text for details of how the ‘mean grades’ were calculated. The χ^2 values indicate the magnitude of the discriminative power of the different survival curves presented. The log-rank test was used as an overall test to compare the Kaplan-Meier survival curves for TG1, TG2 and TG3 penile cancer with respect to CSS. TG1–3 = Tumor grades G1–G3 (Broders’ grade G4 did not occur in the present series).

nitude of interobserver discordance seems to be closely related to the magnitude of the difference in their ‘mean grade’, which obviously reflects inherent ‘aggressive’ versus ‘reserved’ grading characteristics.

Prognostic Impact of the ‘Mean Grade’ per Pathologist on CSS in Uni- and Multivariable Analysis

Median follow-up of the patients alive was 39 months (interquartile range 20–72). CSS rates for the entire study cohort after 1, 3 and 7 years were 91, 70 and 61%. In the univariable analysis, grades assigned by the 7 local pathologists and by the study pathologists A and B significantly predicted CSS (p values < 0.001 ; table 4). Moreover, univariable analyses showed the following parameters to significantly predict CSS: type of surgery, tumor size, tu-

mor thickness (cutoff 5 mm), LVI, tumor stage, pN stage and histologic grade (table 5). The multivariable analyses confirmed an independent prognostic contribution of histopathologic stage and nodal status. Moreover, 3 out of the 5 different multivariable Cox models created disclosed histologic grade, LVI and patient age at surgery to be significantly associated with CSS (table 6). However, the prognostic impact conveyed by the histologic grade was found to be strongly pathologist-dependent (table 6). In particular, the grading rendered by rater B in both rounds conveyed a significant independent prognostic impact on CSS, whereas the grading rendered by rater A only represented an independent prognostic parameter in the second round of examination, while failing to reach statistical significance in the first round. Conversely, the grading rendered by the local pathologists failed to convey any independent prognostic impact in terms of CSS in our multivariable analysis.

Depending upon the incorporation of histologic tumor grade, the different multivariable models showed concordance indices between 0.78 and 0.81 (all p values comparing AUCs were > 0.1). The improvement of the model quality achieved by incorporation of histologic tumor grade ranged between 0.01 and 6.6% (table 6).

Discussion

The histologic grading of surgically treated PSCC might strongly guide clinical management at present. For example, a patient with pT1aG2 PSCC with clinically unremarkable regional nodes will usually be spared ILND, whereas assigning the G3 category to that case would lead to upstaging of the disease (pT1b stage), necessitating

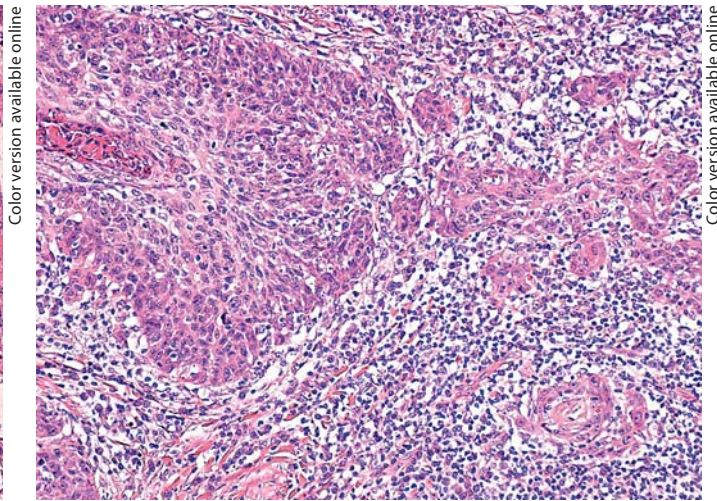
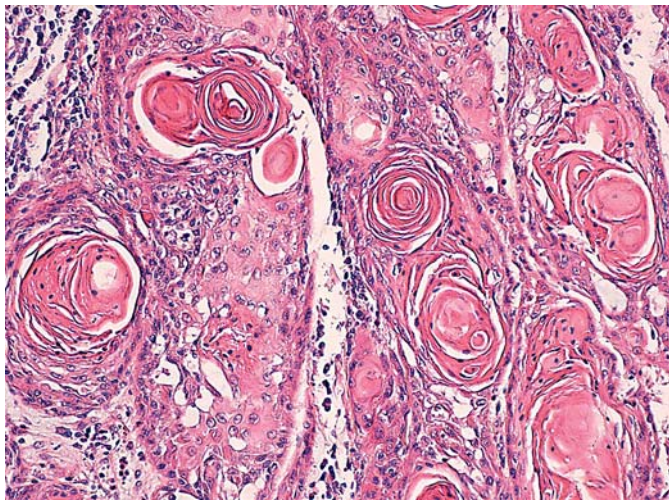


Fig. 1. G1 or G2 squamous cell carcinoma? This study case was graded as G1 by the ‘reserved’ rater B (with the lowest ‘mean grade’) and as G2 by the ‘aggressive’ rater A (with the highest ‘mean grade’). HE. $\times 10$.

Fig. 2. G2 or G3 squamous cell carcinoma? This study case was graded as G2 by the ‘reserved’ rater B (with the lowest ‘mean grade’) and as G3 by the ‘aggressive’ rater A (with the highest ‘mean grade’). HE. $\times 10$.

Table 5. Univariable regression analysis to predict CSS in 147 patients with surgically treated PSCC

Predictor	HR (95% CI)	SE	p value
Age (continuous)	1.00 (0.97–1.03)	0.02	0.824
Total penectomy (referent: partial or excision)	3.32 (1.72–6.44)	0.34	<0.001
Tumor size (continuous)	1.23 (1.01–1.49)	0.09	0.039
Tumor thickness >5 mm (referent: ≤ 5 mm)	3.20 (1.25–8.25)	0.48	0.016
LVI present (referent: absent)	11.38 (4.95–26.17)	0.42	<0.001
pT stage (4 categories)	2.84 (1.94–4.15)	0.19	<0.001
pN+ (referent: pN0/pNx)	9.41 (4.76–18.60)	0.35	<0.001
Year of surgery (continuous)	0.98 (0.89–1.09)	0.08	0.735
Tumor grade			
Local pathology (3 categories)	3.08 (1.78–5.33)	0.28	<0.001
Rater A/1 (3 categories)	2.41 (1.52–3.83)	0.23	<0.001
Rater A/2 (3 categories)	2.65 (1.61–4.34)	0.25	<0.001
Rater B/1 (3 categories)	3.40 (2.16–5.34)	0.23	<0.001
Rater B/2 (3 categories)	3.39 (2.15–5.34)	0.23	<0.001

HR = Hazard ratio.

ILND to be performed, a procedure not to be underestimated in terms of its possible comorbidity [2–9].

We found substantial interobserver discordance, especially between the G2 and G3 categories in pT1 PSCCs (table 2; fig. 2), which continued to be present even after conducting a joint meeting, suggesting inherent ‘aggressive’ versus ‘reserved’ individual grading characteristics of the participating raters.

However, it is likely that conventional histopathology will continue to guide clinical management of PSCC patients in the future since alternative strategies for investigating nodal involvement (e.g. dynamic sentinel node biopsy or ultrasound-guided fine needle aspiration) suffer from false-negative rates, and none of the recently reported prognostic biomarkers has been implemented on a regular basis in clinical practice yet [10, 17–23].

Table 6. Results of the multivariable analyses to evaluate the prognostic impact conveyed by histologic grading based on 5 different Cox models to predict CSS

Predictor	Model 1		Model 2		Model 3		Model 4		Model 5	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Age (continuous)	1.03 (0.99–1.06)	0.052	1.03 (1.00–1.06)	0.050	1.03 (1.00–1.06)	0.047	1.03 (0.99–1.07)	0.034	1.03 (0.99–1.07)	0.031
Total penectomy (referent: partial/excision)	1.31 (0.59–2.89)	0.509	1.34 (0.60–3.02)	0.474	1.35 (0.60–3.06)	0.470	1.09 (0.48–2.48)	0.846	1.07 (0.47–2.45)	0.868
Tumor size (continuous)	0.80 (0.58–1.10)	0.164	0.73 (0.53–1.02)	0.062	0.72 (0.59–1.00)	0.051	0.75 (0.53–1.04)	0.086	0.75 (0.54–1.04)	0.084
Tumor thickness >5 mm (referent: ≤5 mm)	1.26 (0.44–3.67)	0.667	1.12 (0.39–3.27)	0.832	1.05 (0.36–3.09)	0.926	1.11 (0.38–3.22)	0.852	1.11 (0.39–3.32)	0.807
LVI present (referent: absent)	3.61 (1.14–11.4)	0.029	3.32 (1.10–10.0)	0.034	3.25 (1.08–9.83)	0.036	2.72 (0.85–8.77)	0.094	2.75 (0.86–8.83)	0.090
pT stage (4 categories)	2.21 (1.39–3.53)	0.001	2.31 (1.43–3.74)	0.001	2.35 (1.45–3.82)	0.001	2.31 (1.44–3.70)	0.001	2.30 (1.43–3.69)	0.001
pN+ (referent: pN0/pNx)	3.73 (1.42–9.77)	0.007	3.97 (1.50–10.5)	0.006	4.05 (1.52–10.8)	0.005	4.20 (1.51–11.7)	0.006	4.18 (1.51–11.6)	0.006
Year of surgery (continuous)	1.02 (0.92–1.12)	0.644	1.04 (0.92–1.13)	0.526	1.03 (0.91–1.14)	0.582	1.04 (0.92–1.14)	0.485	1.04 (0.92–1.14)	0.492
Tumor grade (3 categories)	1.29 (0.62–2.68)	0.489	1.79 (0.96–3.32)	0.065	2.00 (1.05–3.80)	0.035	2.01 (1.14–3.82)	0.017	2.10 (1.15–3.81)	0.015
Concordance index ¹	0.785 (0.69–0.87)	<0.001	0.798 (0.71–0.88)	<0.001	0.799 (0.71–0.89)	<0.001	0.806 (0.72–0.89)	<0.001	0.804 (0.72–0.89)	<0.001
Model quality (omnibus) ²	+0.01%		+0.87%		+1.73%		+6.38%		+6.60%	

Model 1: incorporation of the 'mean grades' allocated by 7 independent local pathologists; 'mean grade' calculated for study pathologist B in the second round of examination. HR = Hazard ratio.

model 2: incorporation of the 'mean grade' calculated for the study pathologist A in the first round of examination; model 3: incorporation of the 'mean grade' calculated for study pathologist A in the second round of examination; model 4: incorporation of the 'mean grade' calculated for study pathologist B in the first round of examination; model 5: incorporation of the 'mean grade' into the model.

¹p values relate to the comparison between each AUC value and the reference AUC value of 0.5.

²The improvement of model quality in every Cox model employed refers to the comparison of each model's quality with versus without incorporation of the 'mean grade' into the model.

Participation in web-based tutorials might represent a step forward in the standardization of education and training in histopathology [24]. However, regular participation which might be rewarded by the possibility of earning education credit points in order to meet the need of pathology trainees as the principal target group would certainly be necessary in order to prevent participants from reverting to their original form shortly after the sessions are over. The latter phenomenon is likely to have occurred in our study, accounting for substantial interobserver discordance even after conducting a joint meeting. Our findings are in line with the advanced study recently conducted by Naumann et al. [1] on 75 patients with PSCC. Briefly, this study was the first to report on the significant interobserver variation with respect to the grading of penile cancer. However, the magnitude of the prognostic impact entailed by this interobserver discordance was not addressed in that study. In comparison, our data not only demonstrated a great extent of rater dependency obviously attributable to inherent grading characteristics of individual pathologists but, by employing the 'mean grade per pathologist', also showed the magnitude of the prognostic implications linked to this discordance.

There are a few limitations to our study in addition to its retrospective nature. Our evaluation was restricted to the BGS. However, this stipulation seems to be justified since the BGS continues to be the most commonly employed system for grading of PSCCs [25, 26]. Broders' grade 4 did not occur in our series, suggesting that it is a rarity. However, to our knowledge, there are no data on its 'true' incidence available in the literature. Otherwise, the distribution of our grading categories (G1–G3; table 2) is in line with other studies using a three-tiered grading system [2–6, 26, 27]. ILND was performed in only 48 patients (33%) of our study cohort, which precludes reliable evaluation of the impact of histologic grade on nodal involvement. Moreover, a possible change in chemotherapeutic regimens administered during our sampling period cannot be excluded. However, the lack of significance of the hazard ratios for the year of surgery adjusted to all pertinent prognostic parameters strongly indicates that CSS was not influenced by treatment protocols.

In conclusion, conventional histopathology might strongly guide clinical management of PSCC patients at present. The substantial interobserver discordance as well as the dependency of the prognostic impact conveyed by the BGS on the individual pathologist seems, at least in part, to be attributable to inherent grading characteristics of individual pathologists. Therefore, until more reproducible criteria to render the BGS more objec-

tive or novel biomarkers to be determined by the pathologist on a perhaps regular basis become available to guide clinical management, standardization of histopathology training on a perhaps regular basis might be helpful in refining grading of PSCCs.

Acknowledgements

We thank the following department directors for supporting the present study by providing archival wax-embedded tissues for our investigations (shown in alphabetical order): V. Henn, MD

(Department of Pathology, Ruppiner Clinics GmbH, Germany), J. Jander, MD (Department of Pathology, Rhön Clinic Frankfurt/Oder, Germany), H. Lobeck, MD (Department of Pathology, Ernst-von-Bergmann Clinic Potsdam, Germany), R. Pauli, MD (Department of Pathology, Health Care Centre Brandenburg an der Havel GmbH, Germany), and M. Tuffaha (Carl-Thiem Clinic Cottbus, Germany).

Disclosure Statement

The authors declare no competing interests.

References

- 1 Naumann CM, Alkatout I, Hamann MF, et al: Interobserver variation in grading and staging of squamous cell carcinoma of the penis in relation to the clinical outcome. *BJU Int* 2009;103:1660–1665.
- 2 Kattan MW, Ficarra V, Artibani W, et al: GUONE Penile Cancer Project Members: Nomogram predictive of cancer specific survival in patients undergoing partial or total amputation for squamous cell carcinoma of the penis. *J Urol* 2006;175:2103–2108.
- 3 Ficarra V, Zattoni F, Artibani W, et al: Nomogram predictive of pathological inguinal lymph node involvement in patients with squamous cell carcinoma of the penis. *J Urol* 2006;175:1700–1704.
- 4 Zhu Y, Zhang HL, Yao XD, et al: Development and evaluation of a nomogram to predict inguinal lymph node metastasis in patients with penile cancer and clinically negative lymph nodes. *J Urol* 2010;184:539–545.
- 5 Thuret R, Sun M, Abdollah F, et al: Conditional survival predictions after surgery for patients with penile carcinoma. *Cancer* 2011;117:3723–3730.
- 6 Zini L, Cloutier V, Isbarn H, et al: A simple and accurate model for prediction of cancer-specific mortality in patients treated with surgery for primary penile squamous cell carcinoma. *Clin Cancer Res* 2009;15:1013–1018.
- 7 Wittekind C, Meyer HJ (eds): *TNM Classification of Malignant Tumors*, ed 7. New York, Wiley-Blackwell, 2010, pp 225–228.
- 8 Pizzocaro G, Algaba F, Horenblas S, et al: European Association of Urology (EAU) Guidelines Group on Penile Cancer: EAU penile cancer guidelines 2009. *Eur Urol* 2010;57:1002–1012.
- 9 Thuret R, Sun M, Abdollah F, et al: Tumor grade improves the prognostic ability of American Joint Committee on Cancer stage in patients with penile carcinoma. *J Urol* 2011;185:501–507.
- 10 Gunia S, Erbersdobler A, Hakenberg OW, et al: P16(INK4a) is a marker of good prognosis for primary invasive penile squamous cell carcinoma: a multi-institutional study. *J Urol* 2012;187:899–907.
- 11 Cubilla AL, Piris A, Pfannl R, et al: Anatomic levels: important landmarks in penectomy specimens: a detailed anatomic and histologic study based on examination of 44 cases. *Am J Surg Pathol* 2001;25:1091–1094.
- 12 Velazquez EF, Ayala G, Liu H, et al: Histologic grade and perineural invasion are more important than tumour thickness as predictor of nodal metastasis in penile squamous cell carcinoma invading 5 to 10mm. *Am J Surg Pathol* 2008;32:974–979.
- 13 Broders AC: Squamous-cell epithelioma of the skin: a study of 256 cases. *Ann Surg* 1921;73:141–160.
- 14 van Rhijn BW, van Leenders GJ, Ooms BC, et al: The pathologist's mean grade is constant and individualizes the prognostic value of bladder cancer grading. *Eur Urol* 2010;57:1052–1057.
- 15 Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–387.
- 16 DeLong E, DeLong D, Clarke-Pearson D: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
- 17 Leijte JA, Kroon BK, Valdés Olmos RA, et al: Reliability and safety of current dynamic sentinel node biopsy for penile carcinoma. *Eur Urol* 2007;52:170–177.
- 18 Akinde OR, Abudu AK, Anunobi CC, et al: Accuracy of fine needle aspiration in the diagnosis of peripheral lymph node enlargements. Lagos University Teaching Hospital, Nigeria. *Nig Q J Hosp Med* 2011;21:59–63.
- 19 Wu M, Chen H, Zheng X, et al: Evaluation of a scoring system for predicting lymph node malignancy in ultrasound guided fine needle aspiration practice. *Diagn Cytopathol* 2011, E-pub ahead of print.
- 20 Rocha RM, Ignácio JA, Jordán J, et al: A clinical, pathologic, and molecular study of p53 and murine double minute 2 in penile carcinogenesis and its relation to prognosis. *Hum Pathol* 2011;43:481–488.
- 21 Minardi D, d'Anzeo G, Lucarini G, et al: D2-40 immunoreactivity in penile squamous cell carcinoma: a marker of aggressiveness. *Hum Pathol* 2011;42:1596–1602.
- 22 Gunia S, Kakies C, Erbersdobler A, et al: Expression of p53, p21 and cyclin D1 in penile cancer: p53 predicts poor prognosis. *J Clin Pathol* 2012;65:232–236.
- 23 Muneer A, Kayes O, Ahmed HU, et al: Molecular prognostic factors in penile cancer. *World J Urol* 2009;27:1161–1167.
- 24 Kronz JD, Silberman MA, Allsbrook WC: A web-based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by needle biopsy: validation of a new medical education paradigm. *Cancer* 2000;15:1818–1823.
- 25 Velazquez EF, Amin MB, Epstein JI, et al: Members of the Cancer Committee, College of American Pathologists: Protocol for the examination of specimens from patients with carcinoma of the penis. *Arch Pathol Lab Med* 2010;134:923–929.
- 26 Chaux A, Torres J, Pfannl R, et al: Histologic grade in penile squamous cell carcinoma: visual estimation versus digital measurement of proportions of grades, adverse prognosis with any proportion of grade 3 and correlation of a Gleason-like system with nodal metastasis. *Am J Surg Pathol* 2009;33:1042–1048.
- 27 Cubilla AL, Lloveras B, Alejo M, et al: Value of p16(INK)⁴(a) in the pathology of invasive penile squamous cell carcinomas: a report of 202 cases. *Am J Surg Pathol* 2011;35:253–261.