FULL PAPER

BJC

British Journal of Cancer (2016) 114, 1078–1083 | doi: 10.1038/bjc.2016.86

Keywords: Gleason grade; Gleason score; prostate

Validation of a contemporary prostate cancer grading system using prostate cancer death as outcome

Daniel M Berney*[,1], Luis Beltran[1], Gabrielle Fisher[2], Bernard V North[2], David Greenberg[3], Henrik Møller[4], Geraldine Soosay[5], Peter Scardino[6] and Jack Cuzick[2] on behalf of the Transatlantic Prostate Group

[1]Department of Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, EC1A 7BE London, UK; [2]UK Center for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, EC1A 7BE London, UK; [3]National Cancer Registration Service (Eastern Office), Public Health England, CB22 3AD Cambridge, UK; [4]Cancer Epidemiology and Population Health, King's College London, SE1 9RT London, UK; [5]Department of Pathology, Queen's Hospital, Romford, RM7 0AG Essex, UK and [6]Department of Urology, Memorial Sloan-Kettering Cancer Center, New York, 10065 NY, USA

Background: Gleason scoring (GS) has major deficiencies and a novel system of five grade groups (GS$\leqslant$6; 3 + 4; 4 + 3; 8; $\geqslant$9) has been recently agreed and included in the WHO 2016 classification. Although verified in radical prostatectomies using PSA relapse for outcome, it has not been validated using prostate cancer death as an outcome in biopsy series. There is debate whether an 'overall' or 'worst' GS in biopsies series should be used.

Methods: Nine hundred and eighty-eight prostate cancer biopsy cases were identified between 1990 and 2003, and treated conservatively. Diagnosis and grade was assigned to each core as well as an overall grade. Follow-up for prostate cancer death was until 31 December 2012. A log-rank test assessed univariable differences between the five grade groups based on overall and worst grade seen, and using univariable and multivariable Cox proportional hazards. Regression was used to quantify differences in outcome.

Results: Using both 'worst' and 'overall' GS yielded highly significant results on univariate and multivariate analysis with overall GS slightly but insignificantly outperforming worst GS. There was a strong correlation with the five grade groups and prostate cancer death.

Conclusions: This is the largest conservatively treated prostate cancer cohort with long-term follow-up and contemporary assessment of grade. It validates the formation of five grade groups and suggests that the 'worst' grade is a valid prognostic measure.

The Gleason grading of prostate cancer has been established for over 40 years (Gleason, 1966). Although the basic grading categories have remained unchanged in this time, there have been numerous changes in the methodologies used to determine the Gleason score (GS) of prostate cancer over that period.

Changes were first introduced in the 1970s (Gleason and Melliger, 1974). The advent of immunochemistry for basal cell markers in the 1990s introduced further upward shifts in Gleason grading as it was realised that many low-grade lesions diagnosed as prostate cancer were benign lesions such as atypical adenomatous hyperplasia (Bostwick and Chang, 1999; Berney *et al*, 2007). A number of authors cautioned on the diagnosis of very low GSs (Epstein, 2000; Berney, 2007), and this was codified in the 2005 consensus meeting of the International Society of Urological Pathology (ISUP) 2005 (Epstein *et al*, 2005), where it was recommended that scores $<$6 'should rarely if ever' be made.

*Correspondence: Professor DM Berney; E-mail: D.Berney@bartshealth.nhs.uk

Received 15 October 2015; revised 10 March 2016; accepted 10 March 2016; published online 21 April 2016

© 2016 Cancer Research UK. All rights reserved 0007 – 0920/16

Patent Owner Exhibit 2137
IPR2025-01582

More recently, at the 2014 ISUP Chicago conference it was agreed that GSs 2–4 'should not be made' on biopsy (Epstein *et al*, 2016). Although no statements were made concerning GS 5 (3 + 2 or 2 + 3), this score is currently also rarely assigned on biopsy.

There have been further debates since then on unresolved issues on Gleason grading. It has been shown in numerous studies that GS 3 + 3 = 6 tumours show little propensity to recur or metastasise, when completely resected by radical prostatectomy (Miyamoto *et al*, 2009; Ross *et al*, 2012). However, as biopsy specimens remain samples of the tumour, there remains a degree of uncertainty on whether there is un-sampled higher-grade tumour present whenever GS 3 + 3 = 6 is diagnosed.

GS ranges from 2 to 10, but the fact that 6 is the lowest practicable score is very confusing for clinicians and patients (Berney, 2007).

A further concern is that GS 3 + 4 = 7 and GS 4 + 3 = 7 are not separated in most current prognostic tools, although many studies have shown the differences in these scores to be prognostically significant.

In an era when active surveillance is increasingly offered to patients with low-risk prostate cancer, a revision to prostate cancer grading has been proposed (Pierorazio *et al*, 2013) based on five grade groups. This has been accepted by a meeting of senior uropathologists, oncologists and surgeons at an ISUP conference in Chicago in 2014 (Epstein *et al*, 2016). The correlation of GS and grade groups is shown in Table 1.

This grading system has been validated using biochemical relapse as an outcome in a large international series of radical prostatectomy patients (Epstein *et al*, 2015). However, it has not been validated in a conservatively treated cohort, with prostate cancer death at the end point.

There are other crucial refinements in the interpretation of prostate cancer grading, which need to be clarified for use by clinicians and pathologists. There have been some changes to the pattern assignments seen. Cribriform glands and glomeruloid glands, it has been agreed, should all be given a Gleason pattern of four in line with a number of separate lines of evidence on cribriform (Martinez-Rodriguez *et al*, 2007; Dong *et al*, 2013; Kir *et al*, 2014; Kweldam *et al*, 2014; van der Kwast, 2014) and glomeruloid patterns (Pacelli *et al*, 1998; Gobbi *et al*, 1999; Lotan and Epstein, 2009; Liu, Chang *et al*, 2011). There has been debate on whether the 'worst' score seen in a single core of a biopsy series is more or less predictive of outcome than an 'overall' score judged by the pathologist after reviewing the whole series (Kunz and Epstein, 2003; Kunju *et al*, 2009; Tolonen *et al*, 2011). Both 'worst' or 'overall' score are used throughout Europe in pathology practice (Berney *et al*, 2013), although typically the highest score is used by clinicians (Rubin *et al*, 2004).

Table 1. A detailed comparison of contemporary Gleason scoring and grade groups

Comparison of Gleason grading and Grade groups	
GS ⩽ 6	Grade group 1
GS 3 + 4 = 7	Grade group 2
GS 4 + 3 = 7 (if % grade 3 ⩾ 5%)	Grade group 3
GS 4 + 4 = 8 GS 4 + 3 = 7 (if < 5% pattern 3) GS 3 + 5 = 8 GS 5 + 3 = 8	Grade group 4
GS 4 + 5 = 9 GS 5 + 4 = 9 GS 5 + 5 = 10	Grade group 5

Abbreviation: GS = Gleason score.

In this study, we examine the proposed changes in the grading of prostate cancer in a biopsy series treated conservatively and re-reviewed to these new standards. We investigate whether this new grading system can be applied to this data set and whether 'overall' or 'worst' score best predicts prostate cancer death.

MATERIALS AND METHODS

Patients. Cases of prostate cancer were identified from three cancer registries in Great Britain. Within each region, collaborating hospitals were sought and cases from these hospitals were reviewed. Men were included in this study if they were under age 76 years at the date of diagnosis and had clinically localised prostate cancer diagnosed by needle biopsy between 1990 and 2003 inclusively. The median date of diagnosis was May 2002. Patients treated by radical prostatectomy or radiation therapy within 6 months of diagnosis were excluded. In addition, those with objective evidence of metastatic disease (by bone scan, X-ray, radiograph, CT scan, MRI, bone biopsy, lymph node biopsy and pelvic lymph node dissection) or clinical indications of metastatic disease (including pathologic fracture, soft-tissue metastases, spinal compression, or bone pain), or a PSA measurement over 100 ng ml^{-1} at or within 6 months of diagnosis were also excluded. Men who had hormone therapy before the diagnostic biopsy were also excluded, because of the influence of hormone treatment on Gleason pattern. We also excluded men who died within 6 months of diagnosis, or had < 6 months of follow-up.

Original histological specimens from the diagnostic procedure were requested and centrally reviewed by a panel of three expert urological pathologists to confirm the diagnosis of adenocarcinoma and to reassign GSs using of a contemporary and consistent interpretation of the Gleason scoring system (Epstein, 2010). The panel met and discussed all controversial cases and a selection of others to audit the data set. Cribriform and glomeruloid glands were all assigned a Gleason pattern 4. All the cores in each case were given a separate score, and an overall score for the case was also given based on the opinion of the pathologist for each case. Overall grading was assigned by the opinion of the pathologist and the methodology agreed in consensus before analysis. The method chosen was to assign an overall grade thought to be the best estimate of what would be seen at radical prostatectomy. For instance, in a biopsy series with numerous cores with Gleason 4 + 3 = 7 and a small amount of Gleason 4 + 4 = 8 or even higher in a single core, the pathologist might judge that Gleason 4 + 3 = 7 was a more representative score. It was also taken into account that tiny amounts of pattern 5 carcinoma are not included in the grading of radical prostatectomy specimens but given a tertiary score. Percentages of each pattern seen were given. Follow-up was conducted through the cancer registries and the cut-off date was 31 December 2012. Deaths were divided into those from prostate cancer and those from other causes, according to World Health Organisation standardised criteria (WHO, 2010). National ethics approval was obtained from the Northern Multicentre Research Ethics Committee, followed by local ethics committee approval at each of the collaborating hospitals.

Statistical analysis. Survival was analysed with a Cox proportional hazards model. The primary end point was death from prostate cancer. Observations were censored on the date of last follow-up, or at death from other causes. All events were used for estimating hazard ratios (maximum follow-up 232 months), but follow-up was censored at 10 years for predicting 10-year risks. Covariates evaluated were: centrally reviewed overall and worst GS, baseline PSA value, clinical stage, extent of disease (proportion of positive cores), age at diagnosis and use of hormone treatment. Analysis

was repeated substituting 'worst' GS for 'overall' GS and analysed according to the five grade groups.

Baseline PSA concentration was defined as the last pre-diagnostic PSA measurement within 6 months before diagnosis. If no such PSA value was available, we took the first post-diagnostic PSA within 6 months; failing that, the pre-diagnostic PSA taken closest to the date of diagnosis was used. All PSA values after treatment with hormones or orchiectomy or within 3 weeks after a surgical procedure to the prostate were excluded.

PSA concentration was modelled as the natural logarithm of $(1+\text{PSA } (\text{ng ml}^{-1}))$. Patients with values $>100\,\text{ng ml}^{-1}$ were excluded as likely to be metastatic disease. GSs were evaluated in five prognostic grade categories by 'worst' GS and 'overall' GS.

The primary assessment was a univariate analysis of the association between grade group by overall GS and death from prostate cancer and repeated for 'worst' GS. Statistical analyses were done with STATA (version 12, StataCorp, College Station, TX, USA) and R (version 3.0, The R Foundation for Statistical Computing, Vienna, Austria). Multivariate analysis included clinical T stage, diagnostic serum PSA and the volume of disease (percentage of involved cores), and method of treatment (initial hormone treatment or no initial hormonal treatment).

RESULTS

Six thousand five-hundred and one cores from 988 individual cases were assessed for malignancy and graded. The mean, median and interquartile range of patient age, number of cores sampled, serum PSA and percentage of cores involved is shown in Table 2. Cases were divided into the 5 prognostic grade groups from the GS and a comparison between the prognostic grade groups using both 'worst' and 'overall' GS is seen in Figure 1.

Both 'overall' and 'worst' GS analysis yielded highly significant results. The significance of log rank for overall GS in five grade groups was $P=2.79\times10^{-26}$ ($\chi^2=126$ df$=4$). For the worst GS this was $P=1.43\times10^{-24}$ (118 χ^2 df$=4$) with overall GS, therefore, slightly but insignificantly outperforming worst GS. It should be noted that GS $3+4=7$ (grade group 2) separated highly significantly from GS $4+3=7$ (grade group 3). Cox model analysis with hazard ratios by both overall and worst grade group seen also showed high levels of significance (Table 3 and Figure 2). Out of 988 patients, 574 received early hormonal therapy, whereas 414 received watchful waiting only as initial treatment. When analysed separately using overall assessments of grade group for the early hormone-treated group, $P=2.85\times10^{-12}$ ($\chi^2=60$ df$=4$), whereas for the non-hormone-treated group, $P=1.05\times10^{-5}$ ($\chi^2=23.4$ df$=4$).

On multivariate analysis in comparison with log PSA, extent of disease (percentage of involved cores),T stage (stages 3 and 4 merged) and including the method of initial treatment, grade group remained significant with a χ^2 (4df) of 10.3 for overall grade and 9.2 for worst grade. (Table 4), A complete data set was available on 755 patients, with some patients missing details of clinical stage. For the multivariate Cox models, the Harrell c-statistic for overall grade is 0.756 (se$=0.028$) and for worst grade is 0.752 (se$=0.028$).

Removal of extent of disease from the multivariate model, (which was of low significance) resulted in in an increase in log PSA significance with a higher hazard ratio (1.36) and more significant *P*-value (0.010) and on tumour stage 3/4 *vs* 1 with a higher hazard ratio (2.30) and more significant *P*-value (0.010) with similar changes in the worst grade multivariate model; (log PSA hazard ratio$=1.37$, $P=0.008$) and (tumour stage hazard ratio$=2.46$, $P=0.010$).

DISCUSSION

These results show, for the first time, that in a conservatively treated cohort with prostate cancer death as an outcome, interpretation of GS using modern criteria can effectively separate

Table 2. Distribution of the mean, median and interquartile range of patient age, serum PSA, number of cores sampled and percentage of cores involved by tumour across the grade groups

	Overall ($n=988$)	Grade group 1 ($n=307$)	Grade group 2 ($n=303$)	Grade group 3 ($n=210$)	Grade group 4 ($n=56$)	Grade group 5 ($n=112$)
Age (mean (s.d.))	69.6 (5.0)	68.7 (5.4)	69.6 (5.2)	70.3 (4.3)	71.3 (3.4)	69.9 (5.1)
PSA (median (Q1–Q3))	14.2 (8.1–31.0)	8.8 (6.9–13.6)	15.0 (8.6–27.2)	22.9 (11.0–48.5)	19.4 (9.5–48.8)	31.0 (15.9–52.6)
#Cores (median (Q1–Q3))	6 (5–8)	6 (6–8)	6 (5–8)	6 (4–7)	6 (4–8)	6 (4–7)
Tumour % (median (Q1–Q3))	22.0 (7.3–53.4)	6.1 (2.7–12.5)	27.3 (11.5–50.9)	42.4 (18.1–71.4)	33.1 (6.1–68.1)	66.8 (39.9–93.9)

Abbreviation: PSA = prostatic specific antigen.

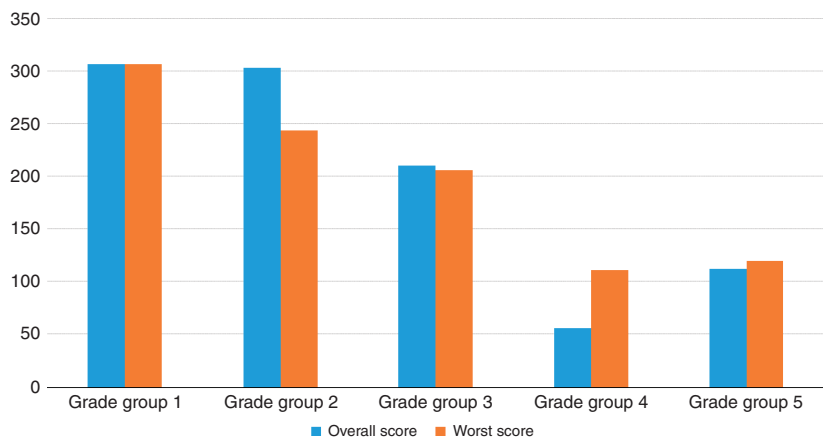

Figure 1. Comparison of overall and worst Grade Group frequencies.

Table 3. Cox Model analysis with hazard ratios by overall and worst grade groups–estimates compared with reference Grade group 1 (GS 3 + 3 = 6)

Grade group (overall)	No. of Men (deaths)	Hazard ratio (95% CI)	*P*-value
1	307 (15)	1 (ref)	
2	303 (39)	2.81 (1.55, 5.10)	0.00067
3	210 (52)	6.05 (3.40, 10.76)	9.5×10^{-10}
4	56 (15)	7.12 (3.48, 14.6)	7.8×10^{-8}
5	112 (48)	12.67 (7.09, 22.64)	$<2 \times 10^{-16}$
Overall $\chi^2 = 110.1$, *P*-value $= 6.9 \times 10^{-23}$.			
Grade group (worst)			
1	307 (15)	1 (ref)	
2	244 (31)	2.69 (1.45, 4.98)	0.0017
3	206 (46)	5.29 (2.94, 9.49)	2.5×10^{-8}
4	111 (28)	6.88 (3.67, 12.9)	1.7×10^{-9}
5	120 (49)	12.0 (6.72, 21.4)	$<2 \times 10^{-16}$
Overall $\chi^2 = 105.9$, *P*-value $= 5.4 \times 10^{-22}$.			

Abbreviations: CI = confidence interval; GS = Gleason score.

five prognostic grade groups. The power of grade groups to predict outcome in this cohort is considerable. It shows that modern interpretation of GS is not only valid using pathological surrogates for outcome or biochemical recurrence but indicates that it correlates with prostate cancer death. We also suggest that grade groups, as suggested in other papers, can be confidently used in reports alongside GS. This will aid both clinicians and patients in their understanding of the severity of the cancer and aid treatment decisions and counselling for active surveillance patients. Gleason scoring presents a 'skewed' scale to patients, with a scale running from 2 to 10, when the lowest valid score is 6. Explaining to patients that a GS 6 cancer is low risk can be difficult. Translation of this to 'grade group 1' will be easier for patients to understand, and for clinicians to explain (Berney, 2007).

It should be especially noted that there is a significant split between GS 3 + 4 = 7 and GS 4 + 3 = 7, (grade groups 2 and 3), which has not been well translated in previous risk assessments such as CAPRA (May *et al*, 2007; Lughezzani *et al*, 2010).

The least significant separation is between GC 4 + 3 = 7 and 4 + 4 = 8, and requires further investigation. Certainly, minor elements of pattern 3 cancer seem to matter little in overall prognosis.

The use of an 'overall' or 'worst' score has been considerably debated in the literature (Kunju *et al*, 2009; Tolonen *et al*, 2011). There is great variability in how GS is assigned in different centres. Some have advocated assigning a GS to every core and giving no 'overall' score for the case. Other pathologists give a GS per submitted specimen pot: which might include more than one core (Berney *et al*, 2013). There have been no direct comparisons of the different methods in a series of conservatively treated prostate carcinomas with long-term outcome. There is a concern that a 'worst' GS might overstate the severity of the disease, especially when the volume of high-grade disease in a single core is small and there is widespread disease of a lower grade in other cores.

We have shown here that the 'worst' GS has a very similar prognostic ability to an 'overall' GS. As it is easier to calculate and relies less on the subjectivity of individual consultant pathologists we advocate its use in routine practice. Using the 'worst' GS, there appears to be greater separation of grade groups 3 and 4. Also the 'worst' GS was used in both the initial and validating studies of grade groups which showed significant differences between grade groups 3 and 4 (Pierorazio *et al*, 2013; Epstein *et al*, 2015). Also, the 'rules' for assigning an overall GS are not clear, and prone to variation between pathologists.

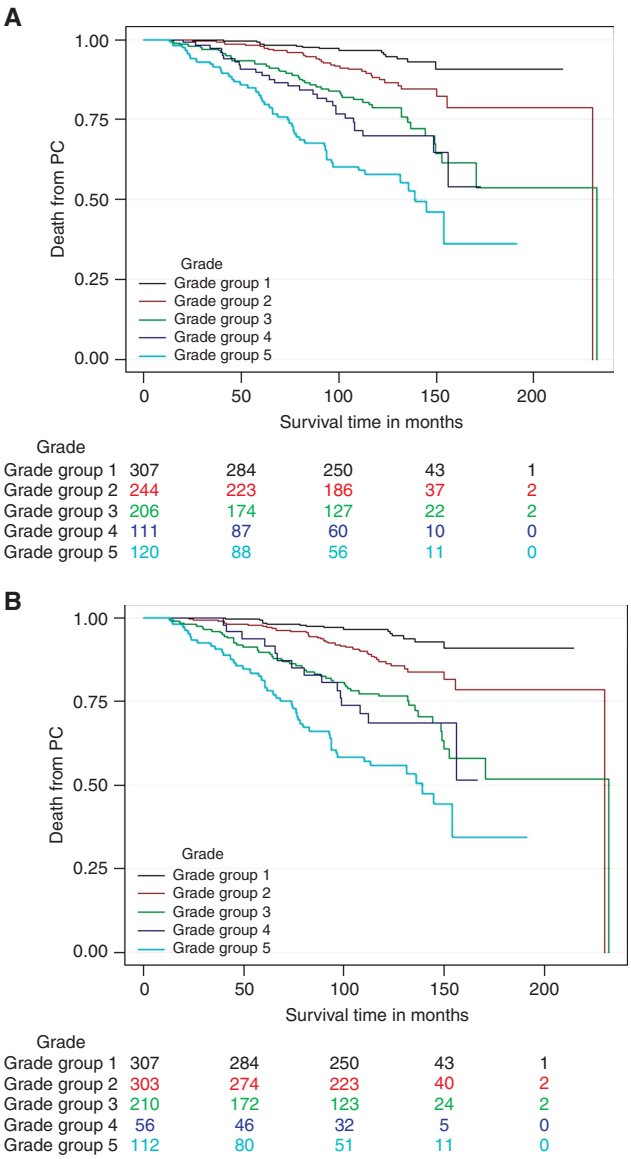

Figure 2. Kaplan–Meier plots of the 5 Grade Groups by worst Grade Group (A) and overall Grade Group (B).

The strengths of this study include the large sample size and detailed nature of the centralised pathological review. In many series it is unclear whether individual cores have been separately graded, especially when they are processed within one cassette or slide.

The weaknesses of the study include its retrospective nature, and the criticism that prostate cancer is no longer treated in the same manner as it was 20 years ago. The majority of the cohort is from sextant biopsies, which is not contemporary practice. This is an unavoidable weakness of current retrospective studies to allow sufficient follow-up to look at prostate cancer death as an outcome. This is a problem for all current long-term studies of prostate cancer outcome, and can also be levelled at large trials such as PROTECT (Oxley *et al*, 2015), where the methods of biopsy are not now standard of care. This will be an on-going problem in prostate cancer outcome studies, with the continuing advance of imaging and template biopsy techniques. For the foreseeable future, pathological grading of prostate cancer will remain standard of care, and adjuvant techniques such as imaging or molecular

Table 4. Multivariate analysis of overall and worst grade group including stage, PSA, initial treatment method and extent of disease. (% disease expressed as % of involved biopsy cores)

Variable	No. of Men (deaths)	Hazard ratio (95% CI)	χ^2 (df)	*P*-Value
Grade group (overall grade)			**9.72 (4)**	**0.046**
1	237 (12)	1 (ref)		
2	237 (32)	1.81 (0.90, 3.66)		0.099
3	159 (41)	2.63 (1.26, 5.47)		0.010
4	43 (9)	2.51 (0.99, 6.28)		0.050
5	79 (28)	3.17 (1.41, 7.12)		0.005
Log PSA		**1.21 (0.95–1.55)**	**3.7 (1)**	**0.122**
T stage			**10.9 (2)**	**0.004**
T stage 1	135 (15)	1 (ref)		
T stage 2	475 (53)	1.04 (0.57, 1.91)		0.894
T stage 3/4	145 (54)	2.11 (1.08, 4.12)		0.028
% disease (per 10%)			**1.66 (1)**	**0.202**
Hormones = *n*	301 (23)	1.05 (0.97, 1.14), 1 (ref)	2.85 (1)	0.100
Hormones = *y*	454 (99)	1.54 (0.92–2.59)		
Grade group (worst grade)			**7.5 (4)**	**0.110**
1	237 (12)	1 (ref)		
2	184 (25)	1.89 (0.92, 3.89)		0.083
3	155 (36)	2.17 (1.03, 4.58)		0.042
4	92 (20)	2.60 (1.19, 5.68)		0.017
5	87 (29)	2.79 (1.24, 6.27)		0.013
Log PSA		**1.21 (0.95–1.55)**	**2.5 (1)**	**0.121**
T stage			**12.1**	**0.0024**
T stage 1	135 (15)	1 (ref)		
T stage 2	475 (53)	1.03 (0.56, 1.90)		0.946
T stage 3/4	145 (54)	2.19 (1.12, 4.26)		0.024
% Disease (per 10%)			**2.05**	**0.153**
Hormones = *n*	301 (23)	1.06 (0.98,1.11), 1 (ref)	3.14 (1)	0.077
Hormones = *y*	454 (99)	1.58 (0.94–2.65)		

Abbreviations: CI = confidence interval; df = degrees of freedom; PSA= prostatic specific antigen.

pathology which are complementary are unlikely to take over from the current gold standard.

In conclusion, we have validated five grade groups in a biopsy series of prostate cancer using prostate cancer death as an outcome. This study compliments other studies using PSA relapse as an outcome for the use of this system internationally.

ACKNOWLEDGEMENTS

We gratefully acknowledge support from Cancer Research UK, ORCHID, a SPORE grant from the US National Cancer Institute (P50CA09629), the David H. Koch Fund and Myriad Genetics. We thank investigators and staff in the cancer registries and participating hospitals for their support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Berney DM (2007) The case for modifying the Gleason grading system. *BJU Int* **100**(4): 725–726.

Berney DM (2007) Low Gleason score prostatic adenocarcinomas are no longer viable entities. *Histopathology* **50**(6): 683–690.

Berney DM, Algaba F, Camparo P, Comperat E, Griffiths D, Kristiansen G, Lopez-Beltran A, Montironi R, Varma M, Egevad L (2013) The reasons behind variation in Gleason grading of prostatic biopsies: areas of agreement and misconception among 266 European pathologists. *Histopathology* **64**(3): 405–411.

Berney DM, Fisher G, Kattan MW, Oliver RT, Moller H, Fearn P, Eastham J, Scardino P, Cuzick J, Reuter VE, Foster CS (2007) Pitfalls in the diagnosis of prostatic cancer: retrospective review of 1791 cases with clinical outcome. *Histopathology* **51**(4): 452–457.

Bostwick DG, Chang L (1999) Overdiagnosis of prostatic adenocarcinoma. *Semin Urol Oncol* **17**(4): 199–205.

Dong F, Yang P, Wang C, Wu S, Xiao Y, McDougal WS, Young RH, Wu CL (2013) Architectural heterogeneity and cribriform pattern predict adverse clinical outcome for Gleason grade 4 prostatic adenocarcinoma. *Am J Surg Pathol* **37**(12): 1855–1861.

Epstein JI (2000) Gleason score 2-4 adenocarcinoma of the prostate on needle biopsy: a diagnosis that should not be made. *Am J Surg Pathol* **24**(4): 477–478.

Epstein JI (2010) An update of the Gleason grading system. *J Urol* **183**(2): 433–440.

Epstein JI, Allsbrook Jr WC, Amin MB, Egevad LL (2005) The 2005 International Society of Urological Pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma. *Am J Surg Pathol* **29**(9): 1228–1242.

Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, Grading C (2016) The 2014 International Society of Urological Pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol* **40**(2): 244–252.

Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, Vickers AJ, Parwani AV, Reuter VE, Fine SW, Eastham JA, Wiklund P, Han M, Reddy CA, Ciezki JP, Nyberg T, Klein EA (2015) A contemporary prostate cancer grading system: a validated alternative to the gleason score. *Eur Urol* **69**(3): 428–435.

Gleason DF (1966) Classification of prostatic carcinomas. *Cancer Chemother Rep* **50**: 125–128.

Gleason DF, Melliger GT (1974) Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* **11**: 7.

Gobbi H, Simpson JF, Jensen RA, Coogan AC, Page DL (1999) Prostatic adenocarcinoma with glomeruloid features. *Hum Pathol* **30**(1): 111–112.

Kir G, Sarbay BC, Gumus E, Topal CS (2014) The association of the cribriform pattern with outcome for prostatic adenocarcinomas. *Pathol Res Pract* **210**(10): 640–644.

Kunju LP, Daignault S, Wei JT, Shah RB (2009) Multiple prostate cancer cores with different Gleason grades submitted in the same specimen container without specific site designation: should each core be assigned an individual Gleason score? *Hum Pathol* **40**(4): 558–564.

Kunz Jr GM, Epstein JI (2003) Should each core with prostate cancer be assigned a separate gleason score? *Hum Pathol* **34**(9): 911–914.

Kweldam CF, Wildhagen MF, Steyerberg EW, Bangma CH, van der Kwast TH, van Leenders GJ (2014) Cribriform growth is highly predictive for postoperative metastasis and disease-specific death in Gleason score 7 prostate cancer. *Mod Pathol* **28**(3): 457–464.

Liu Z, Chang B, Guo SM, Wang T, Li MC, Yang J, Chen K, Yang WM, Liu JH, Ye ZQ (2011) Prostatic adenocarcinoma with glomeruloid structure. *Beijing Da Xue Xue Bao* **43**(4): 519–521.

Lotan TL, Epstein JI (2009) Gleason grading of prostatic adenocarcinoma with glomeruloid features on needle biopsy. *Hum Pathol* **40**(4): 471–477.

Lughezzani G, Budaus L, Isbarn H, Sun M, Perrotte P, Haese A, Chun FK, Schlomm T, Steuber T, Heinzer H, Huland H, Montorsi F, Graefen M, Karakiewicz PI (2010) Head-to-head comparison of the three most commonly used preoperative models for prediction of biochemical recurrence after radical prostatectomy. *Eur Urol* **57**(4): 562–568.

Martinez-Rodriguez M, Ramos D, Mayordomo E, Soriano P, Llombart-Bosch A (2007) Analysis of cribriform Gleason grade 3 in prostatic carcinoma. A histopathological and immunohistochemical study of 32 cases. *Histopathology* **51**(5): 717–719.

May M, Knoll N, Siegsmund M, Fahlenkamp D, Vogler H, Hoschke B, Gralla O (2007) Validity of the CAPRA score to predict biochemical recurrence-free survival after radical prostatectomy. Results from a european multicenter survey of 1,296 patients. *J Urol* **178**(5): 1957–1962, discussion 1962.

Miyamoto H, Hernandez DJ, Epstein JI (2009) A pathological reassessment of organ-confined, Gleason score 6 prostatic adenocarcinomas that progress after radical prostatectomy. *Hum Pathol* **40**(12): 1693–1698.

Oxley J, Simpkin A, Goepel J, Varma M, Griffiths D, Grigor K, Mayer N, Warren A, Deshmukh N, Bhattarai S, Dormer J, Hounsome L, Adamczyk LA, Metcalfe C, Lane JA, Davis M, Donovan JL, Neal DE, Hamdy FC, Robinson MC. T. S. G. Protec (2015) Gleason drift in the NIHR ProtecT study. *Histopathology* **66**(3): 438–446.

Pacelli A, Lopez-Beltran A, Egan AJ, Bostwick DG (1998) Prostatic adenocarcinoma with glomeruloid features. *Hum Pathol* **29**(5): 543–546.

Pierorazio PM, Walsh PC, Partin AW, Epstein JI (2013) Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int* **111**(5): 753–760.

Ross HM, Kryvenko ON, Cowan JE, Simko JP, Wheeler TM, Epstein JI (2012) Do adenocarcinomas of the prostate with Gleason score (GS) ≤ 6 have the potential to metastasize to lymph nodes? *Am J Surg Pathol* **36**(9): 1346–1352.

Rubin MA, Bismar TA, Curtis S, Montie JE (2004) Prostate needle biopsy reporting: how are the surgical members of the Society of Urologic Oncology using pathology reports to guide treatment of prostate cancer patients? *Am J Surg Pathol* **28**(7): 946–952.

Tolonen TT, Kujala PM, Tammela TL, Tuominen VJ, Isola JJ, Visakorpi T (2011) Overall and worst gleason scores are equally good predictors of prostate cancer progression. *BMC Urol* **11**: 21.

van der Kwast T (2014) Re: architectural heterogeneity and cribriform pattern predict adverse clinical outcome for Gleason grade 4 prostatic adenocarcinoma. *Eur Urol* **66**(1): 174.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.