Major shifts in the treatment and prognosis of prostate cancer due to changes in pathological diagnosis and grading

Daniel M. Berney, Gabrielle Fisher*, Michael W. Kattan†, R. Timothy D. Oliver‡, Henrik Møller¶, Paul Fearn§, James Eastham§, Peter Scardino§, Jack Cuzick*, Victor E. Reuter∞ and Christopher S. Foster∞∞ for the Trans-Atlantic Prostate Group

Departments of Histopathology and †Oncology, St Bartholomew’s Hospital, Queen Mary University of London, *Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventative Medicine, London, ∞Department of Cellular Pathology and Molecular Genetics, School of Clinical Laboratory Sciences, University of Liverpool, Liverpool, UK, °°Department of Cellular Pathology and Molecular Genetics, School of Clinical Laboratory Sciences, University of Liverpool, Liverpool, UK, °Departments of Pathology and §Urology, Memorial Sloan Kettering Cancer Center, New York, †Department of Quantitative Health Sciences, The Cleveland Clinic Foundation, Cleveland, Ohio, USA, and ¶Thames Cancer Registry, London, UK

Accepted for publication 1 June 2007

RESULTS

In all, 133 patients (7%) were reassigned a nonmalignant diagnosis. There was a significant reassignment in the Gleason score for those with cancer, with increases of Gleason score across a wide spectrum. In multivariate analysis the revised Gleason score was a more accurate predictor of prognosis than the original score.

CONCLUSION

Misdiagnosis and reassignment of Gleason score at diagnosis would have guided clinicians into large-scale changes in the management of patients. Current rates of misdiagnosis are unknown. If applicable nationally, these changes would have profound effects on the workload of prostate cancer management in the UK.

KEYWORDS

prostate cancer, Gleason grading, misdiagnosis, watchful waiting

INTRODUCTION

The optimum treatment of clinically localized prostate carcinoma is poorly defined, despite 30 142 new cases being reported to cancer registries in the UK in 2001. The choice of treatment for a specific patient is based on various clinical factors [1,2], as well as patient choice [3]. The Gleason score assigned by the pathologist at diagnosis significantly influences the treatment options offered [4]. Accurate pathological diagnosis and grading provides fundamental information to guide patient management [5]. Gleason grading is now considered the ‘reference standard’, providing prognostic information essential for patient assessment at diagnosis [6].

The Trans-Atlantic Prostate Group was established to examine the hypothesis that through a detailed retrospective analysis of outcome in a group of men with clinically localized prostate cancer at diagnosis, variables could be identified that might accurately predict the prognosis [7]. In this section of the study we examined the diagnostic accuracy and concordance with which diagnostic pathologists in the UK applied Gleason grading criteria, together with the likely consequences of the grades assigned to patient management. The objective was to test the hypothesis that the study would reveal a previous broad interpretation of Gleason grading criteria which, when applied to a large group of untreated prostate cancers and standardized according to conventional and internationally agreed criteria [8,9] by a panel of specialist urological pathologists, would be a powerful predictor of prognosis.

PATIENTS AND METHODS

Prostate cancers were examined retrospectively that were localized at
Patients with prostate cancer were identified from six cancer registries in the UK. Within each region, collaborating hospitals were sought and cases from these hospitals were reviewed. Men were included in this study if they were aged <76 years at diagnosis and had clinically localized prostate cancer diagnosed between January 1990 and December 1996. Patients who had a radical prostatectomy or radiation therapy within 6 months of diagnosis, or clear evidence of metastatic disease (by bone scan, X-ray, CT, MRI, bone biopsy, lymph node biopsy or pelvic lymph node dissection) or clinical indications of metastatic disease (including pathological fracture, soft-tissue metastasis, spinal compression or bone pain) at or within 6 months of diagnosis were excluded. Also excluded were men who died within 6 months of diagnosis or had <6 months of follow-up. Eligibility was established by a review of patient records by registry data-collection officers and trained medical staff. Clinical staging was centrally reviewed.

Slides and/or blocks were obtained from the referring hospitals; ≈2% of requested cases were missing or not identifiable in hospital pathology databases, and a further 17% had no original Gleason score given. An initial group of 100 cases, including TURP and needle–biopsy specimens, was examined by the panel of three genitourinary pathologists (D.M.B., V.R. and C.S.F.). This obtained a consensus on the material and agreed criteria on the morphological appearances to be included in the various Gleason patterns [8]; 10% of the remaining cases, where grading was difficult, were reviewed to ensure consistency and standardization in grading criteria. All cases were given a primary and secondary Gleason grade, and the revising panel were unaware of the original Gleason score. In all, 904 (55%) were TURP specimens and 752 (45%) were biopsies.

The outcome was determined through medical record and cancer registry data. Deaths were divided into two categories, death from prostate cancer and death from other causes according to standardized WHO criteria. Patients still alive at the last follow-up were censored at that date. The incremental predictive value of the revised vs original Gleason scores was evaluated by including each Gleason score in a Cox model with other traditional predictors. These two Cox models, one containing the original Gleason score and one the revised Gleason score, were compared in their ability to predict disease-specific survival, with the predictive ability quantified by the concordance index. Similar to the area under a receiver operating characteristic curve, the concordance index measures the probability that the patient who dies from disease first had the lower predicted disease-specific survival in a randomly chosen pair of patients. We used bootstrapping to correct for overfit bias in the concordance index estimation; bootstrapping involves repeated sampling of the dataset and re-estimation of the model to quantify the model accuracy. A Wilcoxon signed-rank test was used to compare the original and reviewed Gleason grades.

### RESULTS

Of 1656 prostate cancer specimens reviewed, 1089 had an original Gleason score assigned; the original and revised Gleason scores are recorded in Table 1, and their distribution is compared in Fig. 1. After the review by the panel, there was a highly significant upgrading of most of specimens (P = 0.001) in both the biopsy and TURP specimens (Fig. 1A,B). Gleason pattern 1 (score 2) was
never assigned by the revising panel but given to >15% during the original diagnosis. Gleason grade 2 (score 4) was rarely assigned by the revising panel, and only in TURP specimens. The original Gleason scores had a much flatter profile, with a relatively even distribution between 2 and 6, and in which few high-grade tumours were diagnosed. Low-score tumours (2–5) were reported in both biopsy and TURP specimens, with slightly more of the lowest grades reported in the TURPs. In 133 (7%) of the referred cases, cancer was not confirmed and contradicted the original pathology report. The survival curve of these patients showed a more favourable outcome than any of the Gleason scored groups, including those with a Gleason score of <6. The misdiagnosis rate showed no variation according to year, and interestingly most of the misdiagnoses were on TURP rather than biopsy material. The material showed either hyperplasia or a spectrum of benign abnormalities, ranging from atypical adenosis to inflammatory atypia.

In a multivariate analysis that adjusted for clinical stage, method of detection, PSA level and age, the original Gleason score was not associated with disease-specific survival (hazard ratio 1.03, 95% CI 0.92–1.15, P = 0.65); the concordance index was 0.682. However, the same Cox model with the revised Gleason score revealed that they were significantly associated with disease-specific survival (hazard ratio 1.71, 95% CI 1.49–1.97, P = 0.001); the concordance index for this model was 0.752 (Fig. 2).

DISCUSSION

This is the largest study yet reported of clinically localized prostate cancer with a PSA follow-up and revised grading of all tumours. It provides a detailed analysis of the natural history of prostatic carcinoma in the UK, together with a unique insight into the routine diagnosis and grading of prostate cancer by general diagnostic surgical pathologists during the period 1990–96. It shows there was a significant error rate both in cancer diagnosis and in Gleason grading.

Accurate Gleason grading is an independent predictor of the likely prognosis, and hence is used as a fundamental piece of information when treatment options are considered. It requires training in the recognition of the basic patterns and constant application to a critical number of new cases to maintain an acceptable level of accuracy in accordance with international consensus [9]. Previous studies confirmed that Gleason grading remains the best single prognostic indicator of the tumour behaviour for an individual patient. It is also highly predictive of recurrence in patients receiving neoadjuvant therapy, and of lymph node and distant metastases occurring in patients receiving no treatment [10–15]. For a system developed over 30 years ago, this is unique and reflects the accuracy with which morphology is a surrogate of biological phenotype. The needle-biopsy Gleason score correlates with the pathological stage at radical prostatectomy [16] and serum PSA level, as well as the expression of many oncogenes. It was shown that the intraobserver variation in Gleason grading is higher when assessed by specialist genitourinary pathologists [17–19].

Many attempts have been made to tighten Gleason’s criteria for individual patterns into various forms of words [20,21]; this has caused considerable revisions in the lower grades of this malignancy. Using strict consensus criteria, no pattern 1 tumours were identified. Pattern 2 tumours were rare or present only in small-volume cancers. During the period 1991–96, there were few specialist genitourinary pathologists in the UK, and hence it is safe to assume that most prostate cancer grading at that time was done by general pathologists who had no specialist interest or training in prostate cancer diagnosis. Current rates of error are likely to be substantially less than in these data from the early 1990s. Since then, education in Gleason scoring has improved greatly, and the use of visual on-line techniques has been highly effective [22] in broadening an understanding of prostate cancer grading. Continued specialization within pathology laboratories will further emphasize this trend in upgrading and should lead to more accurate scoring with enhanced prognostic relevance. A generalized increase in the grade assigned to prostate cancer has been widely acknowledged due to changes in diagnostic criteria [23]. However, the large changes reported here are far greater than any reported previously.

The effect of grade reassignment on the diagnosis and management of prostate cancer is considerable. Low-pattern/score cancers are more likely to be treated conservatively. It is generally accepted that any Gleason score (sum) below 5 (2 + 3 or 3 + 2) indicates indolent disease, where conservative therapy is appropriate. Many now give low Gleason scores, rarely (if ever), especially to biopsy material [24]. Conversely, any enhancement in Gleason grading will be reflected in more patients requiring some form of radical therapy. From these data, it would have resulted in fewer patients considered for ‘watchful waiting’, from about half to <5% of the total. The effect of grade reassignment on routine practice will be just as great for the higher grades of malignancy. The increased frequencies of assigning Gleason score ≥8 might discourage radical surgery.

The correct assignment of individual Gleason scores also has financial implications for patients. The Association of British Insurers suggests that prostate carcinomas are excluded from critical illness cover unless histologically classified as having a Gleason score of ≥6 or having progressed to at least T2N0M0 (http://www.abi.org.uk). Therefore, any enhancement of Gleason score will dramatically affect insurance companies and the number of claimants. Overall, the increasing of Gleason scores to individual prostatic carcinomas will cause more patients to be considered for radical therapy, with an enormous impact on the provision of cancer services throughout the UK. Comparing the survival curves suggests that such grade reassignment is entirely justified by the better prediction of the clinical behaviour of individual cancers.

The misdiagnosis rate of 7% is high; although these patients were not treated aggressively, the potential for patient harm, not least psychologically, remains large. Misdiagnosis was due to recognisable errors in cell identification in most cases. Accurate cancer
diagnosis and Gleason grading is essential for appropriate patient management: Shifts in both the diagnosis and grading methods will have a continued impact on the treatment of prostate cancer throughout the UK and internationally.

AUTHORSHIP CONTRIBUTIONS


ACKNOWLEDGEMENTS

Funding from Cancer Research UK, National Institutes of Health (SPORE) and Koch Foundation. D.M.B. is supported by The Orchid Appeal. The authors’ work was independent of the funders. We gratefully acknowledge the collection of data by cancer registry workers and collaboration from consultants in participating hospitals.

CONFLICT OF INTEREST

None declared.

REFERENCES

4 Gancarzcyk KJ, Wu H, McLeod DG et al. Using the percentage of biopsy cores positive for cancer, pretreatment PSA, and highest biopsy Gleason sum to predict pathologic stage after radical prostatectomy: the Center for Prostate Disease Research nomograms. Urology 2003; 61: 589–95
24 Berney DM. Low Gleason score prostate adenocarcinomas are no longer viable entities. Histopathology 2007; 50: 683–90

EDITORIAL COMMENT

Berney et al. compare the Gleason scores of prostate cancers diagnosed in needle biopsies
and TURP specimens in 1991–1996 with a modern expert review, finding a dramatic increase in grade. Up to 77% of needle biopsies were upgraded by the authors and only 14% were reassigned the original grade. There are many explanations for these changes. Educational efforts clearly play a role. In Europe, prostate cancer was for long graded according to an arbitrary three-tier system, often erroneously referred to as the WHO grading system. When the Gleason system was finally accepted, it quite naturally took some time for pathologists to fully understand it. However, there is also a changed perception of the biological significance of some patterns, even among experts in uropathology. Cribriform and fused patterns are now more likely to be assigned a Gleason pattern 4, rather than a pattern 3. This is reasonable, given the understanding that we now have of the potential aggressiveness of these tumours. In recent years, pathologists have also become increasingly reluctant to use the lower Gleason scores, and most uropathologists now almost never use scores of <4.

Somewhat more problematic is the trend to abrogate Gleason scores 4–5, which in practice reduces the Gleason system to a three-tier system (Gleason scores 6, 7 and 8–10). In this study by Berney et al. no case was assigned a Gleason score 4 or 5 on needle biopsy, and despite their association with transition zone cancers, these scores were even uncommon in TURP specimens. As discussed by the authors, a general upgrading has clinical implications. The prognostic impact of a certain score will be mitigated and comparisons with results from earlier studies will be difficult. Systems for the grading and staging of tumours must be revised continuously as our knowledge of tumour biology develops. However, there are also good reasons to retain a certain stability in these systems. This balance between development and stability requires good judgement from the experts who guide the medical community.

Lars Egevad, Pathology Group, International Agency for Research on Cancer, WHO, Lyon, France

Correspondence: Daniel M. Berney, Department of Histopathology, The Royal London Hospital, 80 Newark St, London E1 2ES, UK. e-mail: D.Berney@bartsandthelondon.nhs.uk, danberney@hotmail.com