Histopathological false-positive diagnoses of prostate cancer in the age of immunohistochemistry.

Luis Beltran MD[1], Amar S Ahmad PhD[2], Holly Sandu MSc[2], Sakunthala Kudahetti MPhil[3], Geraldine Soosay FRCPath[4] , Henrik Møller PhD[5], Jack Cuzick PhD[2], Daniel M Berney FRCPath[1, 3], on behalf of the Transatlantic Prostate Group.

[1] Department of Cellular Pathology, Barts Health NHS Trust, London, UK

[2] UK Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

[3] Department of Molecular Oncology, Barts Cancer Institute Queen Mary University of London, London, UK

[4] Department of Pathology, Queen's Hospital, Romford, Essex, UK

[5] School of Cancer and Pharmaceutical Sciences, King's College London, London, UK.

Correspondence to Luis Beltran

Department of Cellular Pathology, Royal London Hospital, 80 Newark Street, London, E1 2ES, UK. E-mail: luis.beltran2@bartshealth.nhs.uk, Phone: +44(0) 7769115624

Abstract

There are few studies into the rate and causes of histopathological false-positive diagnosis of prostate cancer. Only two of these, including a previous one from our group, incorporate survival data. In addition, in none of the previous studies had immunohistochemistry been originally requested on any of the misdiagnosed cases. Diagnostic biopsies (n=1080) and trans-urethral resection of prostate specimens (n=314) from 1394 men with clinically localised prostate cancer diagnosed in the UK but treated conservatively between 1990 and 2003 were reviewed by a panel of 3 genitourinary pathologists. Thirty five cases were excluded for being potentially incomplete. Of the remaining 1359, 30 (2.2%) were re-assigned to a non-malignant category (26 benign and 4 suspicious for malignancy). Immunohistochemistry had been originally performed on 7 of these. The reasons for the errors were recorded on each case: adenosis (19), partial atrophy (3), PIN (2), seminal vesicle epithelium (1) and hyperplasia (1). Follow-up of these men revealed only one prostate cancer related death, possibly due to un-sampled tumour. In conclusion, a relatively small number of prostate cancer mimics were responsible for a large proportion of the false-positive prostate cancer diagnoses and the use of immunohistochemistry did not prevent the overcall of benign entities as cancer in approximately a quarter of these cases. Targeting these mimics at educational events and raising awareness of the pitfalls in the interpretation of immunohistochemistry in prostate cancer diagnosis, emphasising that glands within a suspicious focus should be treated as a whole rather than individually, may be beneficial in lowering the rate of false-positive diagnosis.

Introduction

Morphological interpretation of H&E stained slides with the aid of immunohistochemistry (IHC) in selected cases remains the gold standard for prostate cancer diagnosis. The increasing incidence of prostate cancer has been fuelled by an aging population and opportunistic Prostatic Specific Antigen (PSA) screening especially in Western developed countries. Mortality is also rising probably also in keeping with an aging population. The vast majority of diagnoses are currently based on needle core biopsies of the prostate, performed according to an extended sextant protocol usually taking 12 cores [1] though template biopsies [2] are increasing. [3,4] This, and the need to re-biopsy men on active surveillance, has caused a substantial increase in the number of cores taken per patient and an extremely heavy burden of diagnostic material for genitourinary pathologists. The reduction in number of cores per patient and number of patients biopsied anticipated with the implementation of targeted biopsies only aiming at areas of the prostate identified as potential cancers on multiparametric MRI [5] has not yet materialised in most institutions.

 The consequences of misdiagnosis of prostate cancer in this modern environment are far reaching for individual patients with both psychological and physical harm possible. Systematic study to try and elucidate the nature and origin of the errors would be highly beneficial with a view to reduce them. However, there have been very few studies into the rate and possible causes of prostate cancer false-positive diagnosis and none have included cases where IHC had been originally performed. [6, 7, 8, 9, 10] A major difficulty in the identification of error includes genuine differences in opinion on borderline cases, making surveys without outcome data of dubious value. Only two studies, [6, 8] including one from our group, incorporate long term follow-up with death from prostate cancer as an outcome. Our previous study was necessarily in a cohort which was predominantly based on trans-urethral resection (TURP) specimens and before IHC had become widespread.

We therefore wished to examine the error rate in a more contemporaneous cohort of conservatively treated prostate cancers with long term follow up in patients who were not offered radical therapy.

Materials and Methods

This work is based on a cohort originally assembled with the aim to study the natural history of prostate cancer. [11] Patients were identified from three cancer registries in Great Britain and collaborating hospitals from these regions were contacted with the request to review the cases. The inclusion criteria were: men under 76 years old at diagnosis with clinically localised prostate cancer diagnosed by biopsy or TURP between 1990 and 2003 inclusively. The median date of diagnosis was May 2002. The exclusion criteria were: radical therapy (surgery or radiotherapy) within 6 months of diagnosis, hormone therapy before biopsy/TURP, metastatic disease recorded by objective evidence (bone scan, X-ray, radiograph, CT scan, MRI, bone biopsy, lymph node biopsy and pelvic lymph node dissection) or by clinical indications (pathologic fracture, soft-tissue metastases, spinal compression, or bone pain), or by a PSA measurement over 100 ng/ml at or within 6 months of diagnosis, death within 6 months of diagnosis and finally, men who had less than 6 months of follow-up. The cohort was thus established in a pragmatic manner on the assumption that 6 months was sufficient to allow for initial treatment options to be decided. Original paraffin-embedded blocks and/or slides used for the primary diagnosis of cancer were requested and centrally reviewed by a panel of three expert urological pathologists (GS, DB and LB) to confirm the diagnosis of adenocarcinoma and to reassign a Gleason score using a contemporary and consistent interpretation of the Gleason scoring system. The panel met and discussed all controversial cases including all the ones where cancer could not be confirmed and a selection of others to audit the data set. All unconfirmed cases were excluded from the primary study on the natural history of prostate cancer and within these, only those that were regarded complete were included in this study, to ensure that the discrepancy was not the result of missing material. Immunohistochemistry using CK5/6 (monoclonal; DAKO, Carpinteria, CA), P63 (monoclonal; DAKO, Carpinteria, CA) and AMACR (monoclonal; DAKO, Carpinteria, CA) was performed if needed. A possible cause for any overdiagnosis was identified by consensus between the

pathologists and recorded in each discordant case. 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.

Results

Out of 1394 cases reviewed consisting of 1080 biopsies (median number of cores=6) and 314 TURPs, a diagnosis of prostate cancer could not be confirmed histologically in 65 specimens. However, in 35 of these, the panel could not be absolutely certain that the material examined was complete, after close examination of block and slide numbers, and these were therefore excluded for the purpose of this study. Of the remaining 30 specimens comprising 2.2% of the cohort, 26 (19 biopsies and 7 TURPs) were re-classified as benign and 4 (all biopsies) were assigned as suspicious for malignancy. The proportion of miss-assigned cases was nearly identical in biopsies (2.1%) and TURPs (2.2%). In five cases the area of concern that had led to the erroneous diagnosis of cancer had been marked by the reporting pathologist on the slide.

The most common error on biopsies was the interpretation of areas of adenosis as cancer (13 of 19 biopsies), followed by partial atrophy, PIN, seminal vesicle epithelium and crowded glands in the context of hyperplasia. Adenosis, including one case of sclerosing adenosis, was the reason for error in 6 of the 7 cases of TURP, whilst no definite reason was noted in one case (Table 1).

IHC for basal cell markers but not for AMACR, due to the cohort's age, had originally been performed on 7 of the miss-assigned cases. It was performed by ourselves, including AMACR, on another 4. Of the 7 cases, 6 contained adenosis and 1 partial atrophy. Basal cell markers demonstrated a discontinuous basal cell layer including the occasional gland completely lacking basal cells in 4 cases.

On the other 2, the technical quality of the staining precluded any meaningful interpretation. In the case of partial atrophy, IHC did demonstrate a basal layer. Examples of the cases are shown in Figures 1-5

Survival data was available for 26 of the 30 discordant cases and their Kaplan-Meier survival curve was compared with that of the prostate cancer cases and their different grades. With 10 years follow-up there was 1 (4%) death from prostate cancer in the 26 cases, which corresponds to a man who had a biopsy triggered by a PSA of 5.6. This death rate was very similar to that of confirmed Gleason 6 cases, found to be 5%. The death rate from Gleason 3+4=7 cases was much higher at 13%. [11]

Discussion

The diagnosis of prostate cancer, particularly on biopsies is challenging, as there are many pathological mimics of malignancy especially where only a limited amount of tissue is seen. There is also no absolute cut off between an atypical small acinar proliferation and malignancy where there is inevitably a degree of subjectivity of diagnosis. [12, 13] Studies examining genuine misdiagnosis in prostate cancer are therefore contentious. The high frequency of the disease in elderly men with a raised PSA and indolent nature of the disease means that false-positive diagnoses will not be appreciated in the majority of men who are not treated surgically. Even after radical prostatectomy, the high incidence of the disease means that small foci of disease seen at radical prostatectomy may not be related to the changes seen at biopsy.

Therefore, there are very few studies into the rate and causes of prostate cancer false-positive diagnosis in histopathology [6, 7, 8, 9, 10] and only two [6, 8] including a previous study by our group, incorporate survival data, being the optimal method to demonstrate indolent behaviour. These studies were based on specimens dating from a time when IHC had not yet been widely implemented; none of the misdiagnosed cases had IHC originally performed. The current study is the largest in terms

of biopsies and the only one including IHC, with 23% of the overcalls having this originally performed. A comparison with the previous studies is shown in Table 2.

We have applied very strict criteria to label a case as false-positive by excluding all re-cuts and potentially incomplete material, which has led to a high exclusion rate. This approach guarantees that the reasons for the false-positive diagnosis recorded are genuine in all cases. However, we are aware that the drawback is that the proportion of false-positives found in the study may be an underestimate. Within this conservative approach, the percentage of false positives stands at 2.2%. The mortality rate of this group was 4% which is very similar to that of Gleason score 3+3=6 (5%) but much lower than that of Gleason score 3+4=7 (13%). [11]

We suggest that the single death from prostate cancer within the false-positive group was most likely due to un-sampled tumour. TRUS biopsy has been found to have a sensitivity of only 48% and a negative predictive value of 63% in detecting clinically significant prostate cancer defined as any Gleason score > or = 3+4 recently. [14] Compared with our previous study, [8] there is a much smaller difference between the mortality rates in the false-positive cases and the Gleason score 6 cases. This is due to the fact that, whilst the mortality in the false-positive group is similar in both studies (6% in the previous and 4% in the current), there is a higher mortality of Gleason score 6 cases in the previous series (13%). This higher mortality can be explained by a number of factors. Firstly, the original series was not based on sextant or decant sampling with the median number of cores being 2, [15] which contrasts with a median of 6 cores in the current study. This under-sampling is more likely to result in under-grading, meaning that many cases of Gleason 3+3=6 would have had un-sampled areas of Gleason pattern 4. Secondly, Gleason grade drift, the consequence of the changes introduced to the Gleason scoring system during the last decade may have led to changes in interpretation. Glomeruloid and small cribriform glands were generally regarded Gleason pattern 3 when the previous cohort was scored. Gleason grade drift has been well documented [16] and is responsible for the Will Rogers phenomenon where any given grade may show better survival over time. [17]

Several benign mimickers of adenocarcinoma have been described [18, 19, 20] of which adenosis, or atypical adenomatous hyperplasia, is the most frequent cause of misdiagnosis in this series (19/30 cases or 63%). Adenosis is mostly present in the transitional zone and therefore is more commonly found in TURP specimens. It consists of a well-defined nodular area of tightly packed small round glands which can display some degree cytological atypia. Its non-infiltrative growth pattern cannot be appreciated in biopsies, where the lesion is typically not fully represented and therefore its recognition is not always straightforward. The basal cell layer is preserved, albeit in a discontinuous fashion focally, in all cases. Depending of the plane of section it is not rare to find one or several acini completely lacking basal cells within the context of otherwise characteristic morphological features of adenosis. [18, 21] It was precisely the finding of occasional acini with no basal cells demonstrated on IHC that appeared to prompt an erroneous diagnosis of malignancy originally in 4 of the cases of adenosis in our series.

Adenosis, or less florid forms of glandular crowding, has also been the most common source of misdiagnosis in previous studies, including our own despite the studies having major differences with regards to the time of original diagnosis, specimen type and clinical context and reasons for review. [6, 7, 8, 9, 10] The first, by Bostwick *et al*, [6] performed on 150 cases originally diagnosed between 1960 and 1970 as low-grade cancer pT1a on TURPs from the Mayo Clinic, found adenosis, including sclerosing adenosis, in 11 (35%) of 35 misdiagnosed cases. Adenosis had not even been described at the time of the original diagnosis and in fact Gleason patterns 1 and 2 were probably based on this entity in many cases. [22] Pathologists were therefore merely following Gleason's descriptions and drawings when diagnosing carcinoma in these cases with no IHC available. On long-term follow up all men with a false-positive diagnosis where free of prostate cancer. A study by Epstein et al [7] was based on biopsies from 535 men originally diagnosed during 1993-1994 and reviewed at Johns Hopkins prior to radical treatment. This found adenosis in 5 (71%) of 7 misdiagnosed cases, in a proportion similar to the current study. No IHC had been performed originally. No follow up was available. Our previous study [8] contained roughly the same number of biopsies and TURP specimens and was based on a cohort

with clinical features identical to the current (men with localised disease conservatively treated during the first 6 months), apart from the period of the original diagnosis (1990-1996) when IHC was not widely available. Adenosis was the cause of the overcall in 38 (28%) of the 133 misdiagnoses. No reason for the discordant diagnosis is specified in the study by Brimo et al. [9]

Contrary to our previous study [8] and that of Bostwick et al [6] no inflammatory conditions were found to be the origin of any overcall in the current study. This may be related to pathologists having a higher threshold to diagnose malignancy in the presence of inflammatory changes.

Herawi et al [23] have looked at benign mimics of prostate cancer from a different perspective. Their study is based on 567 biopsies from consultation cases referred for second opinion where the contributing pathologist questioned malignancy and the expert review classified as benign. No change in diagnosis was made as no firm diagnosis of malignancy was rendered in the first place. Nearly half of their cases had IHC originally performed, which is expected as this cohort is by definition formed by cases deemed to be difficult. Their findings are similar to ours in that patchy or focally absent staining for basal cell markers on IHC in the context of a benign small glandular proliferation raised the possibility of malignancy. Their most common findings were partial atrophy (35.8%) and crowded small glands (25.7%). The difference with our study is that in theirs, the contributing pathologist decided to send the case for expert opinion rather than assigning it to a malignant category. However, no follow up was available.

The issue of false-positive prostate cancer diagnosis has also been examined by Van der Kwast et al, [10] who have analysed data from two of eight centres involved in the European Randomized study of Screening for Prostate Cancer (ERSCP) where all prostate cancer diagnoses performed by general pathologists were routinely screened by experienced reference uropathologists based in each of the centres. The review appears to have happened prior to participants enrolling on the study and therefore it is not clear whether a formal diagnosis of prostate cancer was actually rendered based on any of the re-labelled biopsies. Out of a total of 1950 cases screened, they found 7 (0.36%) where the

diagnosis of prostate cancer was reversed. The most common cause of error was post-atrophic hyperplasia. There is no mention of IHC and no follow up was documented. Their much lower rate of false positives in comparison with the previous published studies and our current work may be related to the fact that their biopsies originate from a very limited number of pathologists from only two centres who have possibly been trained or are being influenced by the reference pathologist who confirmed or refuted their PC diagnoses for trial enrolment purposes.

The observation that IHC had originally been performed on 7 (23%) of the 30 misdiagnosed cases of our study indicates that these had been identified as problematic. However, the result of this test was either misinterpreted or disregarded or in some cases relied upon despite its technical quality being suboptimal. The crucial step of identifying a case as challenging did not prevent an over-call in a relatively high proportion of the discordant cases in this series. Specific training focusing on the interpretation of IHC results in the context of prostate cancer diagnosis could therefore have a major impact in improving the accuracy and reducing the false-positive rate in this field.

If our percentage of false positives is applicable worldwide there are potentially large numbers of men with a misdiagnosis of prostate cancer every year. The majority are probably being labelled as low-grade and nowadays it is reasonable to assume that a large proportion of these men are offered less aggressive forms of therapy or active surveillance. Nevertheless, the clinical, psychological and economic implications are still considerable.

There has been a major decline in the percentage of false-positives compared to our previous study, from 7.5% to 2.2%. [8] That was based on a cohort of men diagnosed during the period 1990-1996 whereas the median date of diagnosis of the current study is 2002. During this time interval, the main changes around prostate cancer histological reporting have been the wide implementation of IHC and subspecialisation in genitourinary pathology. A combination of both factors is probably responsible for the improvement. The percentage of false positives is lower in both papers published by the Johns Hopkins group at 1.1% and 1.2%. [7,9] A possible explanation for this difference lies in the nature of the

cohorts. Theirs is based on cases referred for radical prostatectomy therefore there is selection bias towards more aggressive forms of cancer compared to our cohort formed of men who were treated conservatively. The much higher false-positive rate found by Bostwick *et al* [6] (20%) is hardly surprising taking into account that the original diagnosis was performed at a time when Gleason patterns 1 and 2 were still viable entities. Most of these have subsequently been found to be benign lesions, mainly adenosis. [22]

In conclusion, this is the largest study on the histological false-positive diagnosis of prostate cancer with regards to biopsy numbers and the only one in which IHC was available as a diagnostic tool at the time of the misdiagnosis. We have shown the rate of false-positive cases to be 2.2% and have proved a very favourable outcome on long term follow up for this group, equivalent to that of Gleason 6 and much better than that of Gleason 7. We have identified adenosis as the most common explanation for the diagnostic discrepancy, in concordance with the previous studies on the subject. There has been a major reduction in the proportion of misdiagnoses in comparison with our previous study which we believe is the result of widespread implementation of IHC and specialist reporting in genitourinary pathology. However, 23% of the overcalls appear to be the consequence of misinterpretation of the IHC findings. In particular, it appears that in the context of a suspicious focus of glands with similar morphological features, a proportion of pathologists have treated each individual gland independently when the correct approach is to regard the entire population of glands as a whole at the time of evaluating IHC for basal cell markers. In our opinion specifically targeting interpretation of IHC in training courses and External Quality Assessment schemes could potentially lead to a further decrease in the proportion of false positive histological diagnoses of prostate cancer.

References

1. Philip J, Ragavan N, Desouza J, et al. Effect of peripheral biopsies in maximising early prostate cancer detection in 8-, 10- or 12-core biopsy regimens. *BJU Int*. 2004 Jun;93(9):1218-20.

2. Number of inpatient and day-case SPELLs in England that mention either a transrectal prostate biopsy (M703) or a transperineal prostate biopsy (M702), by year of admission (1 April 2004 to 31 March 2014 (Source: *HSCIC Hospital Episode Statistics*, analysis performed by Public Health England (South West Knowledge and Intelligence Team).

3. Valerio M, Anele C, Charman SC et al. Transperineal template-prostate mapping biopsies: an evaluation of different protocols in the detection of clinically significant prostate cancer. *BJU Int.* 2016 Sep;118(3):384-90

4. Pham KN, Porter CR, Odem-Davis K, et al. Transperineal Template guided prostate biopsy selects candidates for active surveillance – how many cores are enough*? J Urol*. 2015;194:674–679.

5. Moore CM, Robertson NL, Arsanious N, et al. Image-guided prostate biopsy using magnetic resonance imaging-derived targets: a systematic review. *Eur Urol*.2013 Jan;63(1):125-40.

6. Bostwick DG1, Chang L. Overdiagnosis of prostatic adenocarcinoma. *Semin Urol Oncol*. 1999 Nov;17(4):199-205.

7. Epstein JI, Walsh PC, Sanfilippo F. Clinical and cost impact of second-opinion pathology. Review of prostate biopsies prior to radical prostatectomy*. Am J Surg Pathol*.1996 Jul;20(7):851-7.

8. Berney DM, Fisher G, Kattan MW, et al. Pitfalls in the diagnosis of prostatic cancer: retrospective review of 1791 cases with clinical outcome. *Histopathology*. 2007 Oct;51(4):452-7.

9. Brimo F, Schultz L, Epstein JI. The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. *J Urol*. 2010 Jul;184(1):126-30.

10. Van der Kwast TH, Lopes C, Martikainen PMet al. Report of the Pathology Committee: false-positive and false-negative diagnoses of prostate cancer. BJU Int. 2003 Dec;92 Suppl 2:62-5.

11. Berney DM, Beltran L, Fisher G, et al. Validation of a contemporary prostate cancer grading system using prostate cancer death as outcome. Br J Cancer. 2016 May 10;114(10):1078-83.

12. Van der Kwast TH, Evans A, Lockwood G, et al. Variability in diagnostic opinion among pathologists for single small atypical foci in prostate biopsies. *Am J Surg Pathol*. 2010 Feb;34(2):169-77.

13. Epstein JI. Diagnosis of limited adenocarcinoma of the prostate. *Histopathology*. 2012 Jan;60(1):28-40.

14. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. 2017 Feb 25;389(10071):815-822.

15. Rajab R, Fisher G, Kattan MW, et al. Measurements of cancer extent in a conservatively treated prostate cancer biopsy cohort. *Virchows Arch*. 2010 Nov;457(5):547-53.

16. Oxley J, Simpkin A, Goepel J, et al. Gleason drift in the NIHR ProtecT study. *Histopathology*. 2015 Feb;66(3):438-46.

17. Gofrit ON, Zorn KC, Steinberg GD, et al. The Will Rogers phenomenon in urological oncology. *J Urol*. 2008 Jan;179(1):28-33.

18. Trpkov K. Benign mimics of prostatic adenocarcinoma. *Mod Pathol*. 2018 Jan;31(S1):S22-46

19. Gaudin PB, Reuter VE. Benign mimics of prostatic adenocarcinoma on needle biopsy. *Anat Pathol*. 1997;2:111-34.

20. Jones EC, Young RH. The differential diagnosis of prostatic carcinoma. Its distinction from premalignant and pseudocarcinomatous lesions of the prostate gland. *Am J Clin Pathol*. 1994 Jan;101(1):48-64.

21. Bostwick DG, Srigley J, Grignon D, et al. Atypical adenomatous hyperplasia of the prostate: morphologic criteria for its distinction from well-differentiated carcinoma. *Hum Pathol*. 1993 Aug;24(8):819-32.

22. Berney DM. Low Gleason score prostatic adenocarcinomas are no longer viable entities. *Histopathology*. 2007 May;50(6):683-90.

23. Herawi M, Parwani AV, Irie J, et al. Small glandular proliferations on needle biopsies: most common benign mimickers of prostatic adenocarcinoma sent in for expert second opinion. *Am J Surg Pathol*. 2005 Jul;29(7):874-80.

Legends

1. Adenosis. Core biopsy. H&E. Small tightly-packed acini with mild atypia.

2. Adenosis. Core biopsy. 34BE12 immunostain performed originally demonstrates a discontinuous basal cell layer with occasional acini lacking basal cells in this section.

3. Partial atrophy. Core biopsy. H&E. Disorganized acini of different size and shape with no significant atypia and no typical features of atrophy.

4. Partial atrophy. Core biopsy. 34BE12 immunostain performed originally demonstrates a continuous basal cell layer.

5. Adenosis. Core biopsy. H&E. Note fading staining. Dense area combining small and medium-sized acini with no significant atypia.

6. Adenosis. Core biopsy. 34BE12 immunostain performed originally. Staining has failed and appears to highlight nuclei of luminal epithelial cells whilst it is completely negative in other areas.

7. Sclerosing adenosis. TURP. H&E. Small round acini embedded in a richly cellular stroma.

8. Seminal vesicle epithelium. Core biopsy. H&E. Marked nuclear atypia and intracytoplasmic pigment.