Targeted biopsy of the prostate: improvement in detection of high-grade cancer, or the Will Rogers phenomenon?

Edward J Bass MBChB [1,2], Clement Orczyk PhD [1,2], Alistair Grey FRCS (Urol) [1,2], Alex Freeman FRCPath [5], Charles Jameson FRCPath [5], Shonit Punwani FRCR [3,4], Navin Ramachandran FRCR [3,4], Clare Allen FRCR [3,4], Mark Emberton FRCS (Urol) [1,2], Hashim U. Ahmed FRCS (Urol) [6,7]

Affiliations

1. Division of Surgery and Interventional Science, University College London, London, UK

2. Department of Urology, University College London Hospitals NHS Foundation Trust, London, UK

3. Centre for Medical Imaging, Division of Medicine, University College London, London, UK

4. Department of Radiology, University College London Hospitals NHS Foundation Trust, London, UK

5. Department of Histopathology, University College London Hospitals NHS Foundation Trust, London, UK

6. Division of Surgery, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, UK

7. Imperial Urology, Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK

Address for correspondence

Mr Edward Bass, Imperial Urology, Charing Cross Hospital, Imperial College Healthcare NHS Trust, Fulham Palace Road, London, W6 8RF

Email addresses

edward.bass@nhs.net, c.orczyk@ucl.ac.uk, alistair.grey@ucl.ac.uk, alex.freeman@nhs.net, charles.jameson@nhs.net, shonit.punwani@nhs.net, navinramachandran@nhs.net, clare.allen1@nhs.net, m.emberton@ucl.ac.uk, hashim.ahmed@imperial.ac.uk

Key words

Prostatic neoplasms; Epidemiology; Histology; Biopsy

Abstract word count: 231

Manuscript main word count: 2555

Running Head

Is the MRI-targeted prostate biopsy diagnoses more higher Gleason grade cancers. Will this lead to a true improvement in survival? Or create the epidemiological paradox known as the Will Rogers phenomenon?

Abstract

Purpose

The advent of targeted prostate biopsies to suspicious lesions based on imaging confers improved detection of clinically significant prostate cancer. The oversampling of these lesions is likely to better represent the cancer grade. However, such grade inflation might lead to the Will Rogers phenomenon. This study aims to investigate whether patients with Gleason 7 cancer on transrectal biopsy are upgraded following transperineal MRI-targeted biopsy.

Materials and Methods

This retrospective analysis examined 107 consecutive patients presenting at a single tertiary referral centre (July/2012-July/2016) with prostate cancer of Gleason score 7 on transrectal ultrasound-guided systematic non-targeted TRUS-biopsy who then underwent a multi-parametric MRI followed by subsequent visual-estimation MRI-targeted transperineal prostate biopsy for accurate risk stratification and localization. Differences between Gleason grades were compared using Wilcoxon signed-rank tests.

Results

Mean (SD) age was 67.0 (8.0), median (IQR) PSA 6.2 (4.7-9.6) ng/ml. Eighty-four of 107 (78.5%) had Gleason 3+4 on both transrectal systematic biopsy and transperineal MRI-targeted biopsy. Nineteen (17.8%) were upgraded to Gleason 4+3, 3 (3.0%) to Gleason 4+4 and a single (1.0%) patient to Gleason 4+5. These

differences were significant (p=0.0006). Overall, 23/107 (22%) of patients had higher risk disease based on their targeted biopsies (p=0.0006).

Conclusions

There is a significant Gleason grade shift in patients with Gleason 7 prostate cancer on initial TRUS-biopsy who then have transperineal MRI-targeted biopsies. This may suggest a Will Rogers phenomenon through an overestimation of risk.

Introduction

Accurate risk stratification is a cornerstone of modern management of cancer. For prostate cancer it holds a particular prominence, magnified by the variability in its natural history, in terms of potential for progression to lethal, metastatic disease[1]. For example, from the contemporaneous systematic TRUS-biopsy, we would expect to find a disease pattern in a given population, proportions of Gleason 3+3, 3+4 and so forth[2]. In each group, there would be an expected survival given the grading.

Introduction of MRI and targeted biopsies is unlikely to emulate these patterns of grade[3]. Therefore, is the risk of progression conferred by Gleason score on MRI-targeted biopsies the same as that on TRUS-biopsy? Are we shifting those with potentially less aggressive high-grade cancer from a group with classically favourable outcomes, to one with less favourable outcomes? By doing so, the former group becomes less 'contaminated' by patients with higher-grade cancers improving net cancer outcome therein. Likewise, the latter group now has potentially more favourable risk high-grade disease, improving outcomes in this group too.

This so-called Will Rogers phenomenon (WRP) is an epidemiological paradox coined by the American humourist Will Rogers regarding population migration during the Great Depression. Iconic examples are the introductions of novel, more sensitive diagnostic tests, such as CT imaging in lung cancer patients as described by Feinstein et al, detecting previously occult metastases[4]. In prostate cancer, Albertsen first described its effect, postulating that the apparent 28% improvement in prostate cancer outcomes was due to reclassification of Gleason grading in the preceding decade[5]. We aimed to test whether such a phenomenon might be occurring with

the change from random TRUS-biopsies to MRI-targeted biopsies and oversampling of a tumour 'hot-spot' (Figure 2).

Patients and Methods

This retrospective study, performed between July 2012 and July 2016 in a single tertiary referral centre for prostate cancer, evaluated patients presenting for further investigation of prostate cancer after an initial diagnosis on 10-12 core systematic TRUS-biopsy. The inclusion criteria were patients with no palpable malignant disease on DRE (</=cT1c), a Gleason score of 3+4=7 on TRUS-biopsy and a subsequent multiparametric prostate MRI followed by a transperineal visual-estimation MRI-targeted biopsy within six months of the initial TRUS-biopsy. Patients were being evaluated for suitability of focal therapy in which accuracy of risk and location of lesions obligatory and not easily determined on TRUS-biopsy.

MpMRI acquisition was performed according to the European guidelines of Uro-radiology[6]. This includes the use of a 1.5 or 3.0T MRI scanner including T2-weighted axial and coronal, axial DWI and high *b*-value as well as T1-weighted DCE images, utilising intravenous Gadolinium. Each scan was reported by experienced Uro-radiologists and a illustrative diagrammatic map drawn demonstrating regions of interest (ROI) for targeting and scored using a Likert-like scale of 1-5[7]. Scores represented the overall impression of the Uro-radiologist as to the level of suspicion of prostate cancer. Lesions scoring 4 or 5 were regarded as suspicious or highly suspicious for cancer respectively. Those scoring 3 were rated as equivocal. Patients

with lesions scoring 3 or above underwent targeted biopsy. The Likert scoring system was chosen based on the outcome from the 2011 European Consensus Meeting[7] which met prior to the Prostate Imaging and Data Reporting System (PIRADS) multiparametric MRI reporting consensus meeting[8] and has demonstrated equivalence with the PIRADS system[9].

Following mpMRI, visual-estimation targeted biopsy was performed as previously described[10]. Each biopsy was performed under local anaesthetic and sedation in lithotomy position transperineally, under ultrasound guidance with a biplanar 7.5Hz transrectal ultrasound probe (Hitachi Prius).

Histopathological analysis of targeted biopsies was performed at the primary centre by expert Uro-pathologists of over ten years experience each. Pathologists at the primary centre reviewed all the transrectal biopsy slides performed at the referring sites for quality control purposes and to reduce intraobserver variability.

Statistics

Outcomes were primary and secondary Gleason grades as well as Gleason score differences between TRUS-biopsy and MR-targeted biopsy. Significance was tested using the Wilcoxon signed-rank test ($p=0.05$). NCCN risk criteria [11] was also assessed in order to stratify patients before and after MR-targeted biopsies.

Results

Baseline characteristics:

107 consecutive patients were included, all of whom had overall Gleason 3+4=7 cancer in <50% of cores on TRUS-biopsy, followed by transperineal visual-estimation MR-targeted biopsy within median (IQR) 4.7 (3.4-5.2) months. Mean (SD) age was 67.0 (8.0), median (IQR) PSA 6.2 (4.7-9.6) ng/ml. Radiological cancer stage was T2a in 27/107 (25%), T2b in 24/107 (22%), T2c in 45/107 (42%) and T3a in 11/107 (10%).

Outcomes:

Following MR-targeted biopsy, 84/107 (78.5%) had Gleason 3+4=7 disease on both TRUS-biopsy and MR-targeted biopsy. 19 of 107 (17.8%) were upgraded to Gleason 4+3=7, 3 (3.0%) to Gleason 4+4=8 and a single (1.0%) patient to Gleason 4+5=9 (Figure 3) (*p=0.0006*).

In terms of overall cancer risk, TRUS-biopsy compared to MR-targeted biopsy conferred an NCCN favourable-intermediate risk in 107/107 (100%) compared to an NCCN favourable-intermediate risk in 84/107 (79%), unfavourable-intermediate risk in 19/107 (18%), high risk in 2/107 (2%), and very high risk in 2/107 (2%) (p=0.0006) respectively. Overall, 22% of patients had higher risk disease based on their targeted biopsies.

Discussion

These results show that mpMRI-targeted biopsy strategies lead to upgrading of prostate cancer in 1 in 5 patients with Gleason 3+4 prostate cancer after TRUS-biopsy. Further, 1 in 5 patients had a risk shift into less favourable groups.

Our study has limitations. First, although the time between biopsies was low, they were not performed simultaneously and thus, there was a small possibility of disease progression between them. Second, whilst each targeted biopsy was performed in a single centre, the preceding TRUS-biopsy was performed in referring centres. We minimised intra-observer variability of histological reporting by re-reporting these biopsies ourselves. Third, longitudinal cancer control outcomes by which to judge whether the grade and risk shift are clinically significant when matched to a group of patients with Gleason 3+4 disease who did not undergo further biopsies is inexistent.

Common to all oncology, clinicians have advocated aggressive testing to diagnose potentially lethal cancers early and cure them. It is arguable whether such bellicosity has improved in prostate cancer specific survival, since statistical artefacts may be responsible for witnessed reductions in cancer-specific mortality[12]. The introduction of PSA testing in the 1980s led to reductions in death from prostate cancer. Although opinion has swung to-and-fro in regard to the value of PSA as a screening test – indeed, the literature at the time of writing suggests its value is limited[13] – there is no question that early detection led to curable disease being diagnosed, whilst diagnosis with metastasis fell dramatically. However, given the improvements were observed in four years[14], screening may not have responsible.

Instead, epidemiological bias may explain the observation. Both lead-time and length-time bias have been implicated[15].

Further, and pertinent to prostate biopsy, is overdetection. Large proportions of men harbour prostate cancer, even at early ages[16] but most will not die from the disease. PSA screening, and subsequent systematic biopsy overdetects large numbers of these cancers which have little to no impact on life expectancy[17], and, whether treated or not, project an appearance of improved survival outcomes for the cancer as a whole. These epidemiological phenomena are not the WRP, although they can be incorporated within it (Figure 1).

The WRP is a mathematical paradox as demonstrated by the movement of sections of one group to another, leading to an increase in the mean of a value in both groups. When referring to the population migrations in response to the Great Depression of the 1930s, Will Rogers stated 'When the Okies left Oklahoma and moved to California, they increased the average intelligence level in both states'. A comical quip from the humourist, but relevant to medicine where describing an illusory improvement in outcome for groups of patients, but improvement for the individual is absent.

Medically speaking, the phrase was first coined by Feinstein et al in 1985[4]. Reporting on prognoses of patients with lung cancer, they found that a stage migration occurred after introduction of novel diagnostic tests led to the detection of previously undetected synchronous metastases. These occult metastases were more favourable in risk profile compared to those detected by the less accurate, older tests but were now classified as 'unfavourable' risk. Simultaneously, with less

metastatic burden these patients improved the overall cancer outcome in the high-risk group. Likewise, by removing these patients from the favourable prognostic group, there was less overall 'contamination' with metastatic disease and again overall improvement of outcomes for this favourable category. However, no cancer-specific improvements were seen in the entire cohort of patients. Whether there was a difference in outcome granted to each individual by this shift is also arguable. Certainly if the impact of risk shift was to escalate or deny a type of treatment, this could have a profound effect on individuals.

Whilst in other cancers, stage migration is the predominant factor causing the WRP; the dynamics in prostate cancer are more multifaceted. Stage migration *has* occurred, and has done in two ways. First, the introduction of mpMRI has caused an upward stage migration when compared to staging by DRE. For example, Zhang et al found that in a series of 156 clinical T1 cancers undergoing radical prostatectomy, the mpMRI reflected the pathological T stage in around 80% [18]. Thus, the preoperative T-staging is demonstrably improved if mpMRI is used over DRE. However, the implication of doing so in terms of risk stratification is clear. The radiological T3 cases for example, are moved from previously favourable prognostic clinical T1 and T2 groups where previously they will have sat occultly and both groups will demonstrate improved survival. However, in the patient group reported by Zhang et al, all patients received the same treatment, namely radical prostatectomy and one questions whether the individual outcome was altered.

Second, an increase over time in the number of lymph nodes taken at lymph node dissection presents a similar paradox. Undoubtedly, the EPLND offers accurate

nodal staging as the probability of detecting nodal metastasis increases with the number of nodes taken[19]. However, the number of nodes harvested to find one which is positive, correlates inversely to the nodal metastatic burden and thus prognosis. The increase in EPLND in high-risk patients is observable and likely led to the increase in pN1 diagnoses[20]. As with the mpMRI, these patients with low nodal metastatic burden, previously occult, move in to a less favourable nodal metastasis group. This stage migration 'decontaminates' their original group and thus the outcome of both groups improves. Again, the phenomenon treats the individual indifferently.

Traditionally, the WRP and its relationship with prostate cancer manifests itself in migration of grade rather than that of stage. In the 1990s there was consensus that low Gleason grade tumours should be diagnosed rarely. At that time the Gleason grading system recommended that the most and second-most prevalent histological grades determine the score. In 2000 this was amended to allow for when more than two grades are present, and where the worst is not the most or second-most prevalent. In such cases, it was recommended that the most prevalent then the highest grade be reported[21]. Further, Epstein wrote in 2000 that Gleason scores 2-4 should not be diagnosed[22], a concept expanded by consensus to include Gleason score 5. Before 2000, Gleason score 2-5 disease represented up to 50% of new diagnoses[23]. Reported Gleason grades then began to rise and now Gleason scores of 2-5 have disappeared[24].

The consequence of said changes is an adjustment of how patients are characterised as having high-risk disease. In the 1990s most patients were categorised as high-risk

due to their PSA level alone. Indeed, Kane et al reported the proportion as 52%[25]. After the pathological reporting changes in 2000, a high Gleason grade was most commonly responsible for being categorised as such, with Kane et al reporting 65% of high-risk cases being in this manner[25]. In some respects, it is remarkable that risk stratification transitioned from an objective marker - PSA – to one conditional on subjectivity and intraobserver variability. It is not surprising that as this proportion increased, the mean PSA of high-risk disease fell from 30.2ng/mL to 10.4ng/mL [25]. A higher PSA reflects higher disease burden and confers worse outcome. Thus, the effect is the same. The new additions to the high-risk group have a lower disease burden and improve overall survival in the high-risk group, whilst also improving the outcome of the lower-risk groups they were expelled from, due to the new declaration of their small component of high-grade disease. This effect of grade shift on prostate cancer survival was described by Albertsen et al who found the 28% improvement when compared to historical rates was attributable to grade shift as opposed to actual changes in individual outcome[5].

In our own study, targeting an ROI on mpMRI leads to an upgrading of disease from Gleason 3+4 to 4+3 in 17.8%. This is due to higher pathological grade components of disease displaying features that are more readily visible on mpMRI sequences than their lower-grade counterparts. As such, when one targets the ROI, the cores taken retrieve a greater proportion of higher-grade disease. Much like the examples described above, this is another example of grade shift and adds to the causes of the WRP in prostate cancer. For example, the results of the PRECISION trial[26] suggest that 20000 cases of Gleason 4+4 to 5+5 are missed every year by TRUS-biopsy in the

United States[27]. Yet, these cancers are not encountered belatedly with metastases. Indeed, the number of men with metastasis at time of diagnosis continues to fall[28]. One possible explanation is that these cancers with smaller proportions of high-grade disease, detected by our novel diagnostic test, are not behaving in the same manner as those detected by TRUS-biopsy, or at least do not get chance to before said men die from another cause.

This phenomenon will pose a significant challenge. Historical risk stratification systems such as the D'Amico[29], NCCN[11] and MSKCC[30] calculators are not validated for use with targeted biopsy. Instead, reflecting their time of development they are based on TRUS-biopsy. Thus, an upgrading of Gleason score at targeted biopsy if used in these calculators may incorrectly portray an unfavourable prognosis by overestimating the disease risk in these men which may affect the treatment offered. For example, almost one in five of the patients in this study would in all likelihood not have been offered active surveillance for such disease if contemporary risk stratifications are used in the clinical decision making process. However, using those same systems, the counter argument may also be true that had they only undergone TRUS-biopsy, without a subsequent MRI-targeted biopsy, that they might be under-treated. The challenge becomes more urgent when we consider that these calculators are also based on clinical T-stage on DRE. Scepticism therefore should be cast on whether these risk calculators should be applied in patients undergoing either targeted or transperineal mapping biopsies.

Conclusions

There is a significant Gleason grade shift in patients with Gleason 7 prostate cancer on initial TRUS-biopsy who then have transperineal MRI-targeted biopsies. This may suggest a WRP through an overestimation of risk.

Standard Abbreviations

TRUS: Transrectal ultrasound

MRI: Magnetic resonance imaging

CT: Computed tomography

DRE: Digital rectal examination

MpMRI: Multiparametric resonance imaging.

DWI: Diffusion weighted imaging

DCE: Dynamic contrast enhanced

PSA: Prostate specific antigen

EPLND: Extended pelvic lymph node dissection

References

1. Kessler B & Albertsen P. The natural history of prostate cancer. Urol Clin North AM 2003. 30(2):219-26

2. Mahmood U, Levy LB, Nguyen PL, et al. Current clinical presentation and treatment of localized prostate cancer in the United States. *J Urol* 2014. 192(6):1650-6

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

4. Feinstein AR, Sosin DM & Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* 1985. 20;312(25):1604

5. Albertsen PC, Hanley JA, Barrows GH, et at. Prostate cancer and the Will Rogers phenomenon. *J Natl Cancer Inst* 2005. 7;97(17):1248-53

6. Dickinson L, Ahmed HU, Allen C, et al. Magnetic resonance imaging for the detection, localization, and characterization of prostate cancer: recommendations from a European consensus meeting. *Eur Urol* 2011. 59:7477

7. Dickinson L, Ahmed HU, Allen C, et al. Scoring systems used for the interpretation and reporting of multiparametric MRI for prostate cancer detection, localization, and characterization: Could standardization lead to

improved utilization of imaging within the diagnostic pathway? *J Magn Reson Imaging* 2013. **37**:48–58.

8. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. *Eur Radiol* 2012. **22**(4):746-57

9. Rosenkrantz AB, Lim RP, Haghighi, et al. Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate MRI. *AJR Am J Roentgenol* 2013. 201(4):W612-8

10. Bass EJ, Donaldson IA, Freeman A, et al. Magnetic resonance imaging targeted transperineal prostate biopsy: a local anaesthetic approach. *Prostate Cancer Prostatic Dis* 2017. 20(3):311-317

11. Carroll PH & Mohler JL. NCCN Guidelines Updates: Prostate Cancer and Prostate Cancer Early Detection. *J Natl Compr Canc Netw* 2018. 16(5S):620-623

12. Cookson MS. Prostate cancer: screening and early detection. *Cancer Control* 2008. 8:133-140

13. Martin RM, Donovan JL, Turner EL, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. *JAMA* 2018. 319(9):883-895

14. Hankey BF, Feuer EJ, Clegg LX, et al. Cancer surveillance series: interpreting trends in prostate cancer-part1: Evidence of the effects of screening in recent

prostate cancer incidence, mortality, and survival rates. *J Natl Cancer Inst* 1999. 91(12):1017-24

15. Chodak G. Prostate Cancer: Epidemiology, Screening, and Biomarkers. *Rev Urol* 2006. 8(S2):S3-S8

16. Sakr WA, Haas GP, Cassin BF, et al. The frequency of carcinoma and intraepithelial neoplasia in young male patients. *J Urol* 1993. 47:273-287

17. Zlotta AR, Egawa S, Pushkar D, et al. Prevalence of prostate cancer on autopsy: cross-sectional study on unscreened Caucasian and Asian men. *J Natl Cancer Inst* 2013. 105(14):1050-8

18. Zhang J, Hricak H, Shukla-Dave A, et al. Clinical stage T1c prostate cancer: evaluation with endorectal MR imaging and MR spectroscopic imaging. *Radiology* 2009. 253(2):425-34

19. Kim KH, Lim SK, Kim HY, et al. Extended vs standard lymph node dissection in robot-assisted radical prostatectomy for intermediate- or high-risk prostate cancer: a propensity-score-matching analysis. *BJU Int* 2013. 112(2):216-23

20. Budäus L, Spethmann J, Isbarn H, et al. Inverse stage migration in patients undergoing radical prostatectomy: resilts of 8916 European patients treated within the last decade. *BJU Int* 2011. 108:1256-61

21. Gleason DF & Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 1974. 111(1):58-64

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

23. King CR. Patterns of prostate cancer biopsy grading: trends and clinical implications. *Int J Cancer* 2000. 90(6):305-11

24. Ghani HR, Grigor K, Tulloch DN, et al. Trends in reporting Gleason score 1991 to 2001: changes in the pathologist's practice. *Eur Urol* 2005. 24:477

25. Kane CJ, Presti JC Jr, Amling CL, et al. Changing nature of high risk patients undergoing radical prostatectomy. *J Urol* 2007. 177:113

26. Kasivisvanathan V, Ranniko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med* 2018. 378(19):1767-1777

27. Vickers A, Ehdaie B. MRI-Targeted Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med* 2018. 379(6):589

28. Helgstrand JT, Røder MA, Klemann N, Toft BG, Lichtensztajn DY, Brooks JD, et al. Trends in incidence and 5-year mortality in men with newly diagnosed, metastatic prostate cancer-A population-based analysis of 2 national cohorts. *Cancer* 2018. 124(14)(14):2931-2938

29. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998. 280(11):969-74

30. Memorial Sloan Kettering Cancer Center. Prostate Cancer Nomograms. Available from https://www.mskcc.org/nomograms/prostate [Accessed 13th August 2018]

Figure Legends

Figure 1: The incorporation of the epidemiological lead and length-time biases in conjunction with the Will Rogers phenomenon.

Figure 2: This illustrates how the differences in TRUS and MR-targeted biopsy techniques leads to differences in the proportions of different grades of tumour in a given sample being different, even in the same tumour.

Figure 3: This bar chart demonstrates the clear upgrading of 1 in 5 men from Gleason 3+4 to 4+3 and above.