MRI-Targeted Prostate Biopsy Introduces Grade Inflation and Overtreatment

Abderrahim Oussama Batouche\textsuperscript{1,2,3\dagger}, Eugen Czeizler\textsuperscript{1,3\dagger}, Timo-Pekka Lehto\textsuperscript{1,3}, Andrew Erickson\textsuperscript{1,3,4}, Tolou Shadbahr\textsuperscript{1,3}, Teemu Daniel Laajala\textsuperscript{1}, Joona Pohjonen\textsuperscript{1,3}, Andrew Julian Vickers\textsuperscript{5}, Tuomas Mirtti\textsuperscript{1,3,6,7}, Antti Sakari Rannikko\textsuperscript{1,3,8}\ast

\textsuperscript{1}\ast Research Program in Systems Oncology, University of Helsinki, Helsinki, Finland.
\textsuperscript{2}\ast Doctoral program in Computer Science, University of Helsinki, Helsinki, Finland.
\textsuperscript{3}\ast ICAN-Digital Precision Cancer Medicine Flagship, Helsinki, Finland.
\textsuperscript{4}Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK.
\textsuperscript{5}Memorial Sloan Kettering Cancer Center, New York, NY, USA.
\textsuperscript{6}Department of Pathology, Helsinki University Hospital, Helsinki, Finland.
\textsuperscript{7}Department of Biomedical Engineering, Emory University, Atlanta, GA, USA.
\textsuperscript{8}Department of Urology, Helsinki University Hospital, Helsinki, Finland.

\ast Corresponding author(s). E-mail(s): antti.rannikko@hus.fi; Contributing authors: abderrahim.batouche@helsinki.fi; eugen.czeizler@helsinki.fi; timo-peka.lehto@helsinki.fi; andrew.erickson@helsinki.fi; tolou.shadbahr@helsinki.fi; daniel.laajala@helsinki.fi; abderrahim.batouche@helsinki.fi; vickersa@mskcc.org; tuomas.mirtti@hus.fi;
\dagger These authors contributed equally to this work.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Purpose: The use of MRI-targeted biopsies has led to lower detection of Gleason Grade Group 1 (GG1) prostate cancer and increased detection of GG2 disease. Although this finding is generally attributed to improved sensitivity and specificity of MRI for aggressive cancers, it might also be explained by grade inflation. Our objective was to determine the likelihood of definitive treatment and risk of post-treatment recurrence for patients with GG2 cancer diagnosed using targeted biopsies relative to men with GG1 cancer diagnosed using systematic biopsies.

Methods: We performed a retrospective study on a large tertiary centre registry (HUS Acamedic Datalake) to retrieve data on prostate cancer diagnosis, treatment, and cancer recurrence. We included patients with either GG1 with systematic biopsies (3317 men) or GG2 with targeted biopsies (554 men) from 1993 to 2019. We assessed the risk of curative treatment and recurrence after treatment. Kaplan-Meier survival curves were computed to assess treatment- and recurrence-free survival. Cox proportional hazards regression analysis was performed to assess the risk of posttreatment recurrence.

Results: Patients with systematic biopsy detected GG1 cancer had a significantly longer median time-to-treatment (31 months) than those with targeted biopsy detected GG2 cancer (4 months, \( p < 0.0001 \)). The risk of recurrence after curative treatment was similar between groups with the upper bound of 95% CI, excluding an important difference (HR: 0.94, 95% CI [0.71-1.25], \( p = 0.7 \)).

Conclusion: GG2 cancers detected by MRI-targeted biopsy are treated more aggressively than GG1 cancers detected by systematic biopsy, despite having similar oncologic risk. To prevent further overtreatment related to the MRI pathway, treatment guidelines from the pre-MRI era need to be updated to consider changes in the diagnostic pathway.

Keywords: Magnetic Resonance Imaging, Neoplasm Grading, Watchful Waiting, Recurrence, Survival

1 Introduction

Before the early 2010s, PCa was typically diagnosed using systematic transrectal ultrasound (TRUS)-guided biopsy[1]. Recently, MRI-guided targeted biopsies have gained popularity[2]. MRI targeting has been purported to improve both the sensitivity and specificity of biopsy, avoiding biopsy in men at low risk of high-grade cancer, thereby reducing the risk of overdiagnosis and biopsy-related complications [3, 4], and ensuring that regions of the prostate most likely to harbour aggressive disease are sampled. Currently, clinical guidelines, including those of the European Association of Urology, recommend routine use of MRI before diagnostic prostate biopsy.

There is compelling evidence that targeted biopsy has decreased the detection of low-grade cancers and increased the detection of higher-grade cancers compared to systematic biopsy[3, 5–7]. For instance, in the PRECISION trial, men randomised to the MRI pathway had a 13% absolute decrease in Gleason grade group 1 (GG1) cancers and a 12% increase in GG2 cancers[5]. However, it is plausible that this finding is due, at least in part, to the reclassification of GG1 cancers into GG2, that is, grade...
inflation. Indirect evidence for grade inflation comes from the observation that MRI-targeting finds many high-grade cancers in groups of men who are known to have a very low risk of prostate cancer mortality, such as those with negative systematic biopsy[8]. However, there have been few direct comparisons of the relative oncologic risk of tumours identified using MRI-targeted versus systematic biopsy.

In this study, we examine large-scale registry data to compare prostate cancer risks by method of cancer detection. Specifically, we compared the recurrence rates between systematic biopsy-detected GG1 prostate cancer and MRI-targeted biopsy-detected GG2 prostate cancer as a direct evaluation of the grade inflation hypothesis.

2 Materials and methods

We used the Finnprostate dataset, which is a large patient registry study combining Finnish national healthcare data with local hospital data (n=700 000) of men suspected of having PCa (PSA measured) or diagnosed with PCa (Supplementary Fig. 1). From Finnprostate, we gathered a HUS (Hospital District of Helsinki and Uusimaa) sub-cohort of men (n=326 796) with comprehensive patient information regarding outpatient clinic and hospital visits as well as data regarding laboratory tests, medication, radiological, pathological, and surgical reports, as well as comorbidities from 1993 to 2019. The above data is embedded within the regional HUS Academic datalake.

We identified patients with an initial diagnosis of GG1 prostate cancer in systematic biopsies or GG2 prostate cancer in MRI-targeted biopsies. To mitigate the effect of any missing treatment information in our registry on treatments with curative intent, a PSA drop of at least 75% but not less than 3 ng/ml within one year, or at least 50% but not less than 4 ng/ml within one year (this represents 6% of curative treatment data), were considered as an indication of treatment with curative intent, as opposed to active surveillance defined as no curative treatment received or the period from diagnosis until the first curative treatment received. These criteria have been verified to correctly detect 90% of our known first curative treatment data[9]. Depending on the kinetics of the PSA drop, we classified a drop below 0.1ng/ml as radical prostatectomy (RP); otherwise, it was classified as radiation therapy (RT). Sankey diagrams were generated to display the treatment trajectories.

Patients were considered to have experienced recurrence (CR) based on either biochemical recurrence or second-line treatment (Supplementary Fig. 2). Thus, CR considers the clinical reality that some men are being referred to second-line treatment before the official definition of biochemical recurrence has been reached. Biochemical recurrence was defined as a PSA increase over 0.2 ng/ml after RP or 2 ng/ml over the nadir after RT.

2.1 Statistical Methods

Our null hypothesis was that there would be no significant difference in risk for recurrence after treatment between GG1 with systematic biopsy and GG2 with MRI-targeted biopsy. Descriptive statistics were used to summarize and present baseline characteristics, while inferential statistics were used to assess the likelihood of definitive treatment and risk of post-treatment recurrence. Kaplan-Meier survival curves
and log-rank tests were used to analyze treatment- and relapse-free survival. All analyses were performed using R Statistical Software (v4.1.2; R Core Team 2021), including ‘survival’ and ‘survminer’ packages in R software (version 4.1), while Python (version 3.9) was used to pre-process, clean, and combine the data.

We performed a series of sensitivity analyses to assess the robustness of the conclusions. Relapse-free survival was analyzed in patients treated with RP alone to mitigate the effect of androgen deprivation therapy (ADT) with RT may have had. Men who were eventually treated after a period of surveillance longer than one year were more likely to have their PCa evolved over their initial diagnosed GG. Therefore, we separately analyzed a cohort of patients treated within one year of diagnosis, those treated with RP only within one year of diagnosis, and those diagnosed with GG1 and never upgraded in subsequent biopsies. Based on these four sub-cohorts, we performed Cox proportional hazards regression analysis to assess the risk of post-treatment recurrence in GG1 NoMRI and GG2 MRI patients. Furthermore, these analyses were controlled for PSA levels and the number of positive cores.

3 Results

We identified 8407 patients diagnosed with GG1 or GG2 prostate cancer. We excluded patients for whom we were unable to identify their first biopsy with a prostate cancer diagnosis, as well as patients who were over 80 years of age at the time of their first biopsy (Supplementary Fig. 1). The final study cohort consists of 3317 patients with systematic biopsy diagnosed with GG1 cancer and 554 patients with targeted biopsy diagnosed with GG2 cancer (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>GG1 no MRI</th>
<th>GG2 MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>3317</td>
<td>554</td>
</tr>
<tr>
<td>Median age at diagnosis, yr (quartiles)</td>
<td>66 (61-72)</td>
<td>67 (62-72)</td>
</tr>
<tr>
<td>Median PSA at diagnosis, ng/ml (quartiles)</td>
<td>7.3 (5.1-10.5)</td>
<td>8.1 (5.7-12.7)</td>
</tr>
<tr>
<td>Clinical recurrence events (CR in 5yr)</td>
<td>615 (427)</td>
<td>569 (394)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GG1 no MRI</th>
<th>GG2 MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radical prostatectomy</td>
<td>586</td>
<td>165</td>
</tr>
<tr>
<td>Radiation therapy</td>
<td>776</td>
<td>104</td>
</tr>
<tr>
<td>Medical treatment</td>
<td>283</td>
<td>120</td>
</tr>
<tr>
<td>(n=1645)</td>
<td>(n=389)</td>
<td></td>
</tr>
</tbody>
</table>

Treatment patterns were visualised by plotting treatment trajectories in Sankey diagrams for the two groups (Supplementary Fig. 3). In Kaplan-Meier analysis, the likelihood of treatment was significantly lower for men with systematic biopsy detected GG1 compared to men with targeted biopsy detected GG2 HR: 2.77; 95% CI [2.49 – 3.09], p<0.0001) (Figure 1). Risk of treatment within one year of diagnosis was 35% for GG1 and 71% for GG2.
Fig. 1 Five-year treatment-free survival (Kaplan-Meier) for GG1 NoMRI (red line) and GG2 MRI (blue line) groups. 95% confidence interval illustrated with corresponding grey lines. The start time is set as the date of diagnosis. HR: 2.77; 95% CI [2.49 – 3.09], Log-rank p < 0.0001.

Next, we compared the pathology at biopsy with that of prostatectomy in RP-treated men. In 75% of the 268 men who underwent targeted biopsy detected GG2 cancer the biopsy pathology was concordant with RP pathology. In 564 men with systematic biopsy-detected GG1 cancer, the concordance was only 33% (p < 0.0001). In addition, 67% and 16% of the 564 men with GG1 cancer on systematic biopsy and
99% and 24% of the 268 men with GG2 cancer on targeted biopsy, that had radical prostatectomy, had GG2-5 and GG3-5 prostate cancer at final pathology, respectively.

Finally, we compared relapse-free survival between men diagnosed with GG1 cancer using systematic biopsy and men diagnosed with GG2 cancer using targeted biopsy. In the Kaplan-Meier plot, the curves virtually overlapped, with no statistically significant difference in median survival (HR: 0.94, 95% CI [0.71-1.25], p=0.7) (Figure 2).

**Fig. 2** Five-year relapse-free survival (Kaplan-Meier) for GG1 NoMRI (red line) and GG2 MRI (blue line) groups. 95% confidence interval illustrated with corresponding grey lines. The start time is set as the date of treatment. HR=0.94, 95% CI= [0.71-1.25], Log-rank p=0.7.
The results of the sensitivity analyses confirmed the findings of the main analyses. In no analysis was there a statistically significant difference in BCR between systematic GG1 and targeted GG2, and the upper bound of the 95% CI was never greater than 2 (Supplementary Figures 4 – 7 and Supplementary Table 2-3).

4 Discussion

The use of MRI-targeted biopsies has led to a lower detection of GG1 prostate cancer and an increased detection of GG2. Although this finding is generally attributed to the improved sensitivity and specificity of MRI for aggressive cancers, it might also be explained by grade inflation. We found direct evidence to support this hypothesis: GG2 cancers detected by MRI-targeted biopsy were not oncologically more aggressive than GG1 cancers detected by systematic biopsy. Not only were differences between groups non-significant, but the upper bound of the 95% CI also excluded a clinically relevant effect. In our main analysis, the upper bound of the hazard ratio between GG2 and GG1 was 1.25, and the highest upper bound found in the multiple sensitivity analyses was 2. Both values were far lower than those typically reported. For instance, in the MSKCC nomogram, the hazard ratio for BCR comparing GG2 to GG1 was approximately 3[10]. Despite similar oncological risks, MRI-detected GG2 cancers were treated more aggressively, clearly suggesting overtreatment.

In current practice, the major distinction is between GG1, which can be managed conservatively, and GG2 or higher, which generally requires treatment. However, a sub-cohort of low-volume pre-MRI era GG $\leq$ 2 intermediate-risk PCa has been shown to have a similar prognosis to low-risk PCa in AS, according to a recent meta-analysis[11]. Additionally, efforts have been made for more granular risk stratification of intermediate-risk PCa[12, 13]. Furthermore, there is accumulating data on the extent of Gleason pattern 4, total cancer length, PSA density, and perineural invasion determining adverse findings in RP and worse prognosis during follow-up. However, these studies were conducted mainly on systematic biopsy cohorts, and systematic biopsies and targeted biopsies are different and may predict oncologic risk differently[14]. Our results add to the notion that a more precise risk stratification among targeted biopsy-detected cancers is warranted. Hence, before more long-term data accumulate from targeted biopsy studies, treatment decisions should be based on individual risk assessments, not solely on the highest GG stratification and prediction nomograms based on systematic biopsy data[15].

The natural course of the disease is favourable, especially for low- and intermediate-risk PCa, as demonstrated in the recently updated ProtecT trial with only approximately 3% PCa-specific mortality after a 15-year follow-up, irrespective of the treatment[16]. Thus, studies comparing PCa-specific mortality between patients diagnosed with systematic biopsy and targeted biopsy are not expected at any time. Therefore, we considered it justified to use surrogates such as recurrence after curative treatment as an outcome.

The existing literature on targeted biopsy-induced grade inflation is mostly based on data available at diagnosis (pathology at biopsy and radical prostatectomy) and extrapolation into cohorts with virtually no prostate cancer mortality after benign or
low-grade prostate cancer on systematic biopsy, even after an extended follow up [8, 17]. While our results and empirical evidence support grade inflation, they are also in line with both the expert opinion raised in the literature [18, 19] and with the available literature findings [14, 20–22]. Vickers 2021 [8] addressed the possibility of overtreatment related to the use of targeted biopsies. Combining data on the number of men with high-grade cancers detected only on targeted biopsy with data on death rates in men with benign systematic biopsy findings and taking into account current guidelines recommending treatment for high-grade disease, Vickers found that a very large number of men with MRI-detected cancers would need to be treated in order to prevent one death.

Our results support the findings of two previous studies comparing the pathological concordance of systematic and targeted biopsies to radical prostatectomy [20, 21]. Downgrading to GG1 after radical prostatectomy was uncommon regardless of the biopsy method while upgrading was significantly more common in men with GG1 on systematic biopsy. Thus, many men with GG2 on targeted biopsy would often have GG1 on systematic biopsy and are candidates for active surveillance. However, GG1 patients treated with radical prostatectomy likely represent a bias towards a more aggressive cohort of men than GG1 patients, on average.

Indirect evidence for grade inflation also comes from registry studies in which men with negative systematic biopsies have been followed for decades with negligible PCa-specific mortality [23–25]. Furthermore, there is abundant literature showing that up to 30% of men with clinical suspicion of PCa but with negative systematic biopsy de facto harbour high-grade prostate cancer [26]. If these are clinically significant, one would expect PCa mortality during extended follow-up [23].

A clinical implication of our results in the context of the existing literature is that more men with MRI-era GG2 prostate cancer should undergo active surveillance (AS). Ideally, AS protocols should be individualised based on known risk factors, and men should preferably be included in prospective AS trials such as PRIAS [27] or SPCG-17 [28], which should be strongly emphasised in guidelines. In particular, we agree with the statement, “It is not the MRI that is harming by finding indolent grade group 2 cancers that would have been missed by systematic biopsy; the harm is caused by guidelines telling us to treat most of those cancers.” [29, 30].

Our study is prone to some limitations. In particular, one could expect that the subcohort of men with GG1 curatively treated are the ones considered to harbour features of more aggressive disease. Furthermore, a man followed for GG1 cancer may progress to GG2 during follow-up and therefore would be expected to have a similar risk for progression as those with GG2 at diagnosis. Therefore, we performed a series of sensitivity analyses in this study. First, we restricted the analyses to men treated in the first year only to exclude most men who progressed. Next, we excluded men who had progressed on subsequent biopsies from the analyses. Finally, we adjusted the risk of recurrence for PSA level and the number of positive cores as covariates in the multivariable Cox regression model. These sensitivity analyses confirmed our original findings, with no statistically significant differences between groups, and the upper bound of all 95% CI was far below published estimates of the hazard ratio for GG2 vs. GG1 disease. Due to the registry nature of our study, some of the data were
granular, and some of the missing data were inferred (e.g., curative treatment based on PSA changes). This could have caused inaccuracies in the results. However, the effect should be limited based on a quality assessment of the data. Clinical indications for MRI and curative treatments are not known, and indications likely have changed during the study period, which represents a limitation, although it is unclear if this would cause a bias. The Gleason grading also changed during the study period. This has caused risk inflation, especially for GG1, with a bias favouring the GG1 group, further strengthening our results and interpretation. The strengths of our study lie in the large sample size and the comprehensive clinical data available for each patient from diagnosis to treatment and follow-up. Our registry included patients from the beginning of the MRI era, allowing for a comparatively long follow-up period. This extended follow-up period provides valuable insights into the long-term effects of treatments or interventions and helps us better understand outcomes over time.

5 Conclusion

GG2 cancers detected by MRI-targeted biopsy are treated more aggressively than GG1 cancers detected by systematic biopsy, despite having similar oncologic risks. To prevent further overtreatment related to the MRI pathway, treatment guidelines from the pre-MRI era need to be updated to consider changes in the diagnostic pathway.

Financial disclosures

Andrew Vickers receives royalties from the sale of the 4Kscore.

Funding/Support and Role of the Sponsor

This work was supported by grants from the Cancer Society Finland (Tuomas Mirtti and Antti Rannikko), Academy of Finland (Tuomas Mirtti), Sigrid Jusélius Foundation (Tuomas Mirtti), Jane and Aatos Erkko Foundation (Antti Rannikko) and state funding for university-level health research (Tuomas Mirtti and Antti Rannikko). This work was also supported in part by the National Institutes of Health/National Cancer Institute (NIH/NCI) with a Cancer Center Support Grant to Memorial Sloan Kettering Cancer Center [P30-CA008748], a SPORE grant in Prostate Cancer to Dr. H. Scher [P50-CA92629], a PCORI grant [ME-2018C2-13253], the Sidney Kimmel Center for Prostate and Urologic Cancers, and David H. Koch through the Prostate Cancer Foundation. The funding sources did not play any role in the study design, execution, data analyses, writing of the report, or submission of the article for publication.

References


[18] Next generation imaging (NGI) in prostate cancer: scan or not to scan? That is the question. Technical report, Nordisk Urologisk Förening (Scandinavian Association of Urology), Helsinki, Finland (June 2022)


