Clinical translation of VERDICT MRI biomarkers
for Prostate Cancer characterisation

Saurabh Singh
University College London

PhD. Thesis, Submitted for the degree of Doctor
of Philosophy, University College London, 2021
DECLARATION

I, Saurabh Singh, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

14/11/2021
Abstract

Prostate Cancer is the second most common cancer worldwide and the second most common cause of cancer death among men in the UK. The introduction of multiparametric magnetic resonance imaging (mpMRI) in the diagnostic pathway in the last 10 years has helped reduce the over-diagnosis of clinically insignificant prostate cancer whilst still detecting clinically significant cancer (csPCa). However, mpMRI is not a perfect test and has important shortcomings. The specificity of mpMRI is modest at 37% due to indeterminate appearances caused by benign pathologies. This causes men to undergo unnecessary invasive biopsies that are negative for cancer. Therefore, there is a need for biomarkers that can aid the assessment of MRI for prostate cancer.

In this thesis, I evaluate whether quantitative imaging biomarkers (QIBs) derived from VERDICT MRI (Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumours) can add more value in classifying which men have csPCa when compared to biomarkers derived from mpMRI in a cohort of men suspected to have prostate cancer.

I first analyse the performance of the current diagnostic pathway in a prospective cohort and identify in which participants there is a need for additional biomarkers. I then compare the diagnostic accuracy of VERDICT QIBs to mpMRI QIBs in differentiating men with csPCa to those without in men who underwent biopsy in the prospective cohort. I also assess the image quality and inter-reader agreement of the most promising VERDICT parameter map when qualitatively assessed by radiologists.

Finally, I identify whether there is a clinical need for biomarkers in monitoring response to hormonal and radiotherapy in patients with locally aggressive prostate cancer. Subsequently, I study the changes in VERDICT QIBs after hormonal and radiotherapy in men with biopsy-proven prostate cancer.
Acknowledgements

This body of work has only been possible because of collaboration and the support of several individuals. I would like to acknowledge a few of them here.

I would like to thank my family for their constant love and support; my wife Swetha, my mum Usha, my dad Satyendra, my father-in-law Mrutyunjaya Rao, my mother-in-law Santi, my sister Prachi, my brothers-in-law Ashish and Bhargav, my niece Saanvi and my nephew Vedaant.

I would like to thank my supervisor Shonit Punwani for giving me an opportunity and funding me to work on this project. He has been an inspiring and supportive supervisor who has been very generous with his time and attention. I am always amazed at how efficient and balanced he is whilst being involved in so many projects. I aspire to develop this poise in my academic and clinical career.

I would like to thank Dr Steve Bandula for allowing me to start this journey in 2018.

I am grateful to Harriet Rogers for her help and advice on managing a large dataset. I would like to thank Baris Kanber for his help with XNAT.

Daniel Alexander, Laura Panagiotaki, Thomy Mertzanidou and the CMIC team have been great collaborators and I am thankful for their help.

Joey Clemente and Teresita Beeston have been a joy to work with and extremely helpful in running the trial.

Louise Dickinson and Francesco Giganti have been generous with their time reading prostate MRIs.

I am also thankful to the deanery and Royal College of Radiologists for allowing me to take three years out of programme to pursue a PhD.

I would also like to thank every patient for their generosity in taking part in this research.

“Prajnanam Brahma” “Tat tvam asi”.

(Knowledge is all, that thou art).
Impact Statement

This work represents a forward step in translating QIBs which have shown potential in the lab to the clinic, with the aim of better risk stratifying men who have a clinical suspicion of prostate cancer.

Specifically, I have identified a group of men where qualitative mpMRI assessment leads to a high number of false positives and where VERDICT QIBs could have an impact. I have shown that quantitative assessment of VERDICT biomarkers could potentially reduce false positives and subsequent unnecessary biopsies in men who go on to have negative or insignificant cancer detected on histology. I have derived thresholds based on data analysed in this thesis, which I plan to validate in a prospective trial where biopsy decisions would be made by the application of VERDICT MRI.

I have also shown that image quality and inter-reader agreement is similar for fractional intracellular volume (FIC) maps compared to apparent diffusion coefficient maps. This finding is important in justifying that VERDICT MRI could replace ADC maps in clinical protocols in the future. Furthermore, I have shown that FIC maps can be assessed by radiologists qualitatively as well as quantitatively. Results showed that a similar number of true positives were detected by both mpMRI and VERDICT datasets but there are a lower number of false positives with FIC maps. This finding is important for the future prospective trial design. It suggests that if biopsy decisions are made from assessing FIC maps that this would not lead to a significant number of missed cancers.

Finally, I identified a clinical need for better biomarkers to detect the failure of hormonal and radiotherapy in men with locally aggressive prostate cancer. I found that a significant proportion of men (22-33%) have residual or recurrent disease two years after therapy. This leads to poor oncological outcomes. I assessed the changes that can be seen in mpMRI QIBs and VERDICT QIBs in a small number of men undergoing hormonal and radiotherapy. I found that VERDICT may have a role in monitoring therapy after EBRT and the timing of imaging needs at least 6 months after therapy. The next step will be to study men who have disease
recurrrence and investigate whether VERDICT QIBs can help detect them earlier than mpMRI.

**Academic Output**

**Peer-reviewed Journal Articles relevant to this thesis:**


Other peer-reviewed Journal articles published during PhD time:


Selected Peer-reviewed Conference Abstracts relevant to thesis:

Singh S, Rogers H, Kanber B, et al. An opportunity for men with positive prostate mpMRI studies to safely avoid biopsy – Results from the INNOVATE Study – European Association of Urology Conference July 2021

Valindria V, Palombo M, Chiou E, Singh S, Punwani S, Panagiotaki E. Synthetic Q-Space Learning with Deep Regression Networks for Prostate Cancer Characterisation with VERDICT IEEE 18th International Symposium on Biomedical Imaging April 2021

Valindria V, Singh S, Palombo M et al. Non-invasive Gleason Score Classification with VERDICT-MRI International Society of Magnetic Resonance Medicine March 2021

Singh S, Rogers H, Johnston E, et al. VERDICT MRI fractional intracellular volume assessment could help avoid unnecessary biopsies in men assessed for prostate cancer with multi-parametric MRI European Congress of Radiology, Vienna, July 2020

Palombo M, Singh S, Whitaker H et al. Relaxed-VERDICT: decoupling relaxation and diffusion for comprehensive microstructure characterization of prostate cancer. ISMRM August 2020
Contents

ABSTRACT ...........................................................................................................................................3

ACKNOWLEDGEMENTS ..................................................................................................................4

This body of work has only been possible because of collaboration and the support of several individuals. I would like to acknowledge a few of them here. ..........4

IMPACT STATEMENT .......................................................................................................................5

ACADEMIC OUTPUT ........................................................................................................................6

CONTENTS ........................................................................................................................................8

LIST OF FIGURES ............................................................................................................................14

LIST OF TABLES ................................................................................................................................17

LIST OF ABBREVIATIONS ................................................................................................................18

1 OVERVIEW ......................................................................................................................................21

1.1 PROBLEM STATEMENT .............................................................................................................21

1.2 CHAPTER SUMMARY ................................................................................................................21

2 INTRODUCTION ............................................................................................................................24

2.1 PROSTATE CANCER – THE DISEASE ......................................................................................24

2.1.1 Epidemiology .........................................................................................................................24

2.1.2 Aetiology ................................................................................................................................24

2.1.3 Pathogenesis ..........................................................................................................................25

2.1.4 Clinical Features ....................................................................................................................26

2.1.5 Investigations .........................................................................................................................26

2.2 THE ROLE OF MRI IN THE ASSESSMENT OF PROSTATE CANCER ..................................28

2.2.1 mpMRI and Biopsy ................................................................................................................31

2.3 HISTOPATHOLOGY ......................................................................................................................32

2.4 MANAGEMENT ............................................................................................................................35

2.5 NOVEL IMAGING BIOMARKERS (IB) DEVELOPMENT AND ASSESSMENT ..................36
2.5.1 Technical Validation ................................................................. 37
2.5.2 Biological and Clinical Validation ............................................. 37
2.5.3 Cost-effectiveness ................................................................... 38
2.5.4 VERDICT biomarker development .......................................... 39

3 INTRODUCTION TO MAGNETIC RESONANCE IMAGING .................. 43
3.1 THE NUCLEAR MAGNETIC RESONANCE SIGNAL ............................. 43
3.1.1 Net magnetisation .................................................................... 44
3.1.2 Excitation ................................................................................ 44
3.1.3 Relaxation ................................................................................ 46
3.2 PULSE SEQUENCES ..................................................................... 49
3.2.1 Free Induction Decay ............................................................... 49
3.2.2 Spatial Localisation ................................................................. 50
3.2.3 Diffusion Weighted Imaging .................................................... 56
3.2.4 Microstructural models and VERDICT ...................................... 63

4 CLINICAL OUTCOMES FROM THE INNOVATE STUDY ................... 67
4.1 INTRODUCTION .......................................................................... 67
4.1.1 Aims ...................................................................................... 69
4.1.2 Null Hypotheses ..................................................................... 69
4.2 METHODS .................................................................................. 70
4.2.1 Study Design .......................................................................... 70
4.2.2 Study population ..................................................................... 70
4.2.3 Multiparametric MRI protocol .................................................. 71
4.2.4 VERDICT MRI Protocol ........................................................... 74
4.2.5 Biopsy .................................................................................... 74
4.2.6 Serum PSA and PSA Density ................................................... 74
5.3.4  PSA Density .................................................................................................................... 99
5.3.5  ADC ................................................................................................................................ 101
5.3.6  FIC .................................................................................................................................... 106
5.3.7  FEES .................................................................................................................................. 107
5.3.8  FVASC .............................................................................................................................. 109
5.3.9  Receiver Operator Curve Analysis (ROC) .......................................................................... 111

5.4  DISCUSSION .......................................................................................................................... 113

6  INTER-READER AGREEMENT, IMAGE QUALITY AND QUALITATIVE ASSESSMENT OF VERDICT MRI IN A MULTI-READER STUDY ......................................................................................................................... 117

6.1  INTRODUCTION ..................................................................................................................... 117

6.1.1  Aims .................................................................................................................................... 119

6.1.2  Null Hypotheses ................................................................................................................ 119

6.2  METHODS .............................................................................................................................. 121

6.2.1  Participants ......................................................................................................................... 121

6.2.2  Reader training ................................................................................................................... 122

6.2.3  IMAGE QUALITY .............................................................................................................. 122

6.2.4  Likert and Pirads SCORING .............................................................................................. 125

6.2.5  Time taken ......................................................................................................................... 125

6.2.6  Reference Standard ............................................................................................................ 125

6.2.7  Statistical Analysis ............................................................................................................. 127

6.3  RESULTS ............................................................................................................................... 130

6.3.1  Participants ......................................................................................................................... 130

6.3.2  Image Quality .................................................................................................................... 130

6.3.3  Inter-reader AGREEMENT ................................................................................................. 133

6.3.4  Reading time ...................................................................................................................... 133
9.1.1  Clinical OUTCOMES OF THE INNOVATE TRIAL .................................................. 182
9.1.2  Meta-analysis of biopsy outcomes in patients treated with ebrt .......................... 217
9.1.3  PREVALNCE OF PROSTATE CALCIFICATION in MEN WITH PROSTATE CANCER........... 241
9.1.4  Selected Conference Abstracts ........................................................................... 256
List of Figures

Figure 2-1 Original diagram showing Gleason grade and expected morphology by Dr Donald Gleason [41]................................................................. 33
Figure 2-2 Current stage of biomarker development for VERDICT MRI.......... 40
Figure 3-1 Relaxation T1 is the time for net magnetisation (Mz) to recover to (1 - 1/e) or 63.2% ......................................................................................... 47
Figure 3-2 Transverse relaxation (Mxy) – T2 is time taken for Mxy to decay by 1/e or 37% ............................................................................. Error! Bookmark not defined.
Figure 3-3 Free Induction Decay and Fourier transformation (FT)............... 50
Figure 3-4 Inversion Recovery for fat suppression.................................. 55
Figure 3-5 The diffusion sensitive spin echo......................................... 57
Figure 3-6 Echo Planar Pulse Sequence ................................................. 58
Figure 3-7 K-space trajectory for echo-planar imaging (left) compared to the conventional cartesian trajectory (right) .............................................. 59
Figure 3-8 Signal decay curve showing the deviation from the gaussian mono-exponential assumed decay of the signal............................................. 61
Figure 4-1 Study patient selection ........................................................... 71
Figure 4-2 Proportion of patients biopsied for each Likert score............. 77
Figure 4-3 Proportion of patients with clinically significant cancer subdivided by Likert score ........................................................................... 77
Figure 4-4 MRI images of 4 participants ................................................ 78
Figure 4-5 Boxplot of PSA subdivided by Likert score and decision to biopsy..... 80
Figure 4-6 Distribution of PSA subdivided by Likert score and presence of Clinically significant cancer on biopsy Outliers are indicated by (-) ............... 81
Figure 4-7 Distribution of PSA Density (PSAD) by Likert score and decision to biopsy...................................................................................... 83
Figure 4-8 Distribution of PSA Density (PSAD) by Likert score and presence of Clinically Significant Cancer................................................................. 83
Figure 4-9 PSA density (PSAD) and Gleason score, Outliers are indicated by (-) 84
Figure 4-10 Proportion of patients with clinically significant cancer by PIRADS score ..................................................................................................... 85
Figure 4-11 Distribution of PSA and PIRADS score. Outliers are indicated by (-). ............................................................................................... 85
Figure 4-12 Distribution of PSAD and presence of clinically significant cancer... 86
Figure 5-1 MR Images from three men in the INNOVATE study......................... 97
Figure 5-2 Distribution of PSA density by Likert and PIRADS score and presence of clinically significant cancer................................................................. 99
Figure 5-3 Distribution of PSA Density and Gleason grade.............................. 100
Figure 5-4 Distribution of lesion-ADC subdivided by Likert and PIRADS Score and presence of clinically significant cancer............................................................ 102
Figure 5-5 Lesion-ADC and Gleason score correlation. ................................ 102
Figure 5-6 Lesion ADC distribution subdivide by Likert score and presence of Clinically Significant cancer for men who underwent mpMRI at 3T. Statistical significance is indicated by ‘*’. Outliers are denoted by (-) ...................................... 104
Figure 5-7 Lesion ADC distribution and Gleason grade for men who underwent mpMRI at 3T........................................................................................................... 104
Figure 5-8 Lesion ADC distribution subdivide by Likert score and presence of Clinically Significant cancer for men who underwent mpMRI at 1.5T............ 105
Figure 5-9 Lesion ADC distribution and Gleason grade for men who underwent mpMRI at 3T........................................................................................................... 105
Figure 5-10 Distribution of FIC by Likert and PIRADS score and presence of clinically significant cancer............................................................. 106
Figure 5-11 Lesion FIC and Gleason Grade correlation.................................... 107
Figure 5-12 Distribution of lesion-FEES subdivided by Likert and PIRADS Score and presence of clinically significant cancer............................................................ 108
Figure 5-13 Lesion FEES and Gleason grade correlation.................................. 109
Figure 5-14 Distribution of lesion-FVASC subdivided by Likert and PIRADS Score and presence of clinically significant cancer............................................................ 110
Figure 5-15 Lesion FVASC and Gleason grade............................................... 111
Figure 5-16 ROC curves for FIC, PSAD, ADC, FEES, FVASC...Error! Bookmark not defined.
Figure 5-17 ROC curves for FIC, PSAD, ADC, FEES, FVASC............................ 112
Figure 6-1 PIQUAL Scoring system................................................................. 120
Figure 6-2 Study Design................................................................................. 121
Figure 6-3 Locked Sequential Read scheme ................................................. 123
Figure 6-4 ADC maps (left) and FIC maps (right) from 5 participants........... 124
Figure 6-5 Reference Standard ...................................................................... 127
Figure 6-6 Reporting pro-forma for the Qualitative study............................... 129
Figure 6-7 Likert Quality Scores for both readers; FIC on left, ADC on right... 132
Figure 6-8 PIQUAL Quality scores for both readers........................................ 132
Figure 7-1 Study Schedule.................................................................................. 148
Figure 7-2 MRI Images of a 78-year-old man with a Gleason 3+4 tumour in the anterior left prostate............................................................... 150
Figure 7-3 Trend of PSA and PSA density for 5 patients................................. 152
Figure 7-4 Trend of apparent diffusion coefficient (ADC) for 5 patients......... 154
Figure 7-5 Trend of fractional intracellular volume (FIC) for 5 patients.......... 156
Figure 7-6 Trend of fractional extravascular extracellular volume (FEES) for 5 patients.................................................................................................... 158
Figure 7-7 Trend of fractional vascular fraction (FVASC) for 5 patients......... 160
Figure 7-8 Change in gland volume at the four defined time points.............. 161
Figure 8-1 Imaging Biomarker roadmap for VERDICT MRI after this Thesis.. 167
List of Tables

Table 2-1 Comparison of Likert and Prostate Imaging Reporting and Data System Scoring ........................................................................................................................................ 30
Table 4-1 INNOVATE Scan Parameters.................................................................................................................. 73
Table 5-1 VERDICT scan parameters................................................................................................................................ 94
Table 6-1 Application of reference standard to study cohort.................................................................................. 136
Table 6-2 Application of overall reference standard to Reader scores (+) denotes entire dataset including T2W, FIC or ADC, high b-value image and DCE images. .................................................................................................................................................. 137
Table 6-3 Application of sub-group reference standard to Reader scores (+) denotes entire dataset including T2W, FIC or ADC, high b-value image and DCE images.................................................................................................................................................. 139
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Apparent Diffusion Coefficient</td>
</tr>
<tr>
<td>ADT</td>
<td>Androgen Deprivation Therapy</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>BPH</td>
<td>Benign Prostatic Hyperplasia</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CMI</td>
<td>Centre for Medical Imaging</td>
</tr>
<tr>
<td>CNR</td>
<td>Contrast-to-noise ratio</td>
</tr>
<tr>
<td>CoV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>CZ</td>
<td>Central Zone</td>
</tr>
<tr>
<td>DCE</td>
<td>Dynamic Contrast Enhanced</td>
</tr>
<tr>
<td>DICOM</td>
<td>Digital Imaging and Communications in Medicine</td>
</tr>
<tr>
<td>DKI</td>
<td>Diffusion Kurtosis Imaging</td>
</tr>
<tr>
<td>DRE</td>
<td>Digital Rectal Examination</td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion tensor imaging</td>
</tr>
<tr>
<td>DUS</td>
<td>Distal urethral sphincter</td>
</tr>
<tr>
<td>DWI</td>
<td>Diffusion Weighted Imaging</td>
</tr>
<tr>
<td>EES</td>
<td>Extravascular, Extracellular Space</td>
</tr>
<tr>
<td>EPI</td>
<td>Echo Planar Imaging</td>
</tr>
<tr>
<td>FEES</td>
<td>Fractional Extracellular Extravascular Space</td>
</tr>
<tr>
<td>FIC</td>
<td>Fractional Intracellular Volume</td>
</tr>
<tr>
<td>FID</td>
<td>Free Induction Decay</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>FVASC</td>
<td>Fractional Vascular Fraction</td>
</tr>
<tr>
<td>GBCA</td>
<td>Gadolinium Based Contrast Agent</td>
</tr>
<tr>
<td>GE</td>
<td>Gradient Echo</td>
</tr>
<tr>
<td>GV</td>
<td>Gland Volume</td>
</tr>
<tr>
<td>H&amp;E</td>
<td>Haematoxylin and Eosin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>INNOVATE</td>
<td>Combining advances in imaging with biomarkers for improved diagnosis of aggressive prostate cancer</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile Range</td>
</tr>
<tr>
<td>IRAS</td>
<td>Integrated Research Approval System</td>
</tr>
<tr>
<td>ISM RM</td>
<td>International Society for Magnetic Resonance in Medicine</td>
</tr>
<tr>
<td>IVIM</td>
<td>Intravoxel Incoherent Motion</td>
</tr>
<tr>
<td>mp</td>
<td>Multiparametric</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic resonance</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NICE</td>
<td>The National Institute for Health and Care Excellence</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PACS</td>
<td>Picture Archiving and Communication System</td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>PI-RADS</td>
<td>Prostate Imaging Reporting and Data System</td>
</tr>
<tr>
<td>PSA</td>
<td>Prostate Specific Antigen</td>
</tr>
<tr>
<td>PSAD</td>
<td>Prostate-Specific Antigen Density</td>
</tr>
<tr>
<td>PSMA</td>
<td>Prostate-Specific Membrane Antigen</td>
</tr>
<tr>
<td>PZ</td>
<td>Peripheral Zone</td>
</tr>
<tr>
<td>QIB</td>
<td>Quantitative Imaging Biomarker</td>
</tr>
<tr>
<td>QIBA</td>
<td>Quantitative Imaging Biomarkers Alliance</td>
</tr>
<tr>
<td>REC</td>
<td>NHS Research and Ethics Committee</td>
</tr>
<tr>
<td>RF</td>
<td>Radiofrequency</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Spin Echo</td>
</tr>
<tr>
<td>SI</td>
<td>Signal Intensity</td>
</tr>
<tr>
<td>SMART</td>
<td>Serial mp-MRI scanning in prostate cancer after Androgen deprivation therapy and RadioTherapy</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>T2W</td>
<td>T2-weighted</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TE</td>
<td>Time to Echo</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>TPM</td>
<td>Transperineal Template Mapping Biopsy</td>
</tr>
<tr>
<td>TR</td>
<td>Time to Repetition</td>
</tr>
<tr>
<td>TRUS</td>
<td>Transrectal Ultrasound</td>
</tr>
<tr>
<td>TSE</td>
<td>Turbo Spin Echo</td>
</tr>
<tr>
<td>TV</td>
<td>Tumour Volume</td>
</tr>
<tr>
<td>TZ</td>
<td>Transition Zone</td>
</tr>
<tr>
<td>UCL</td>
<td>University College London</td>
</tr>
<tr>
<td>UCLH</td>
<td>University College London Hospital</td>
</tr>
<tr>
<td>VERDICT</td>
<td>Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumours</td>
</tr>
</tbody>
</table>
1 Overview

1.1 Problem Statement

Multiparametric MRI has transformed the diagnosis of prostate cancer (PCa) in the last 10 years and is now the first-line investigation for men suspected of localised prostate cancer. With increasing use, there is an opportunity to further improve prostate cancer characterisation and better risk stratify men suspected of PCa. Quantitative imaging biomarkers are an active area of research in imaging but few if any are used in clinical practice. My thesis aims to determine whether there is a need for biomarkers in prostate cancer diagnosis, evaluate which biomarkers from VERDICT MRI could add value and assess how they can be used clinically to aid clinical decisions.

1.2 Chapter Summary

Chapter 2 – Background, Literature Review and Framework

The first part of this introductory chapter provides an overview of prostate cancer as a disease and a summary of what’s known about epidemiology, clinical presentation, diagnostic tests, and clinical management. The performance of diagnostic tests is evaluated from the published literature, with particular focus on multiparametric MRI and PSA which are currently used. I then establish a framework for evaluating a biomarker in the context of prostate cancer at the end of the chapter.

Chapter 3 – MRI Physics

This chapter outlines the fundamentals of the different components of a multiparametric MRI with a focus on diffusion-weighted imaging. I compare the different models available for diffusion modelling and explore why novel imaging biomarkers from VERDICT (Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumours) could improve prostate cancer assessment.
Chapter 4 – Clinical Outcomes from INNOVATE

In this chapter, I evaluate the clinical outcomes from a prospective observational study of men suspected of prostate cancer. I analyse the performance of mpMRI and PSA in risk stratifying men for biopsy. I identify which men could benefit from further risk stratification.

Chapter 5 – VERDICT MRI Quantitative Analysis

In chapter 5, I analyse the distribution of VERDICT biomarkers in men who have clinically significant cancer compared to those without from the INNOVATE cohort who underwent biopsy. I compare VERDICT biomarkers to mpMRI biomarkers in differentiating csPCa and establish thresholds from receiver operating curve analysis.

Chapter 6 – VERDICT MRI Qualitative Analysis – Multi-reader study

Two radiologists with similar experience in reading mpMRI evaluated imaging datasets including fractional intracellular volume (FIC) maps from VERDICT using a locked sequential model for image quality and prostate cancer risk. I compare the image quality of FIC maps to ADC maps and assess inter-reader agreement. I also explore if the qualitative assessment of individual FIC maps could reduce false positives compared to ADC maps.

Chapter 7 - The effect of Androgen deprivation therapy and Radiotherapy on VERDICT MRI parameters

I first establish if there is a need for an imaging biomarker in monitoring therapy after hormonal and radiotherapy for prostate cancer by conducting a meta-analysis on positive biopsy rates after treatment. I then study the changes in the prostate due to hormonal therapy and radiotherapy as seen on multiparametric MRI and VERDICT MRI. As part of the ‘Serial mp-MRI scanning in prostate cancer after
androgen deprivation therapy and RadioTherapy’ (SMART), patients with prostate cancer underwent VERDICT MRI before ADT, after ADT but before the start of radiotherapy, during radiotherapy and after radiotherapy. I will analyse the change in VERDICT parameters to understand the effects of ADT and radiotherapy on the prostate.

Chapter 8 – Thesis Conclusions and Future Outlook

The concluding chapter provides a summary of the findings and potential for future work in translating VERDICT MRI in clinical practice.
2 Introduction

2.1 Prostate Cancer – The Disease

2.1.1 Epidemiology

Prostate Cancer (PCa) is the second most common cancer in men worldwide and the second most common cause of cancer death among men in the UK. From the 1990s onwards, the incidence has increased by 41% in UK men [1]. The peak rate of prostate cancer cases is in older men aged between 75 and 79 [1].

Despite substantial research, prostate cancer risk factors have not been conclusively identified. Higher rates of prostate cancer are seen with increasing age and in men of Afro-Caribbean origin and with family history. There is no strong evidence for any preventable risk factors. According to the international agency for research on Cancer/World Cancer Research Fund classifications, there is only limited or probable evidence of increased risk with androgenic steroid use, red meat ingestion and body fat content [2].

Survival with prostate cancer is high but depends on whether it is detected at an early stage. For instance, 100% of people with prostate cancer survive their disease for 5 years if diagnosed at the earliest stage compared to 49% if diagnosed at its latest stage (Stage IV) [3].

2.1.2 Aetiology

The aetiology of prostate cancer remains not well understood. Some risk factors are thought to play a role in pathogenesis, but the evidence is not causal. Hormones such as androgens and insulin-like growth factor have been associated[4]. Support for this lies in androgen deprivation therapy for prostate cancer. However, the role of androgens may be permissive because androgens are required to maintain
prostatic epithelium. From a genetic perspective, as in other cancers, prostate cancer develops through the accumulation of genetic and epigenetic changes, resulting in inactivation of tumour suppressor genes, activation of oncogenes and dysregulation of housekeeping genes\[5\] [6]. Putative links have been made with inflammation and the role of the microbiome which establish a favourable microenvironment for prostate cancer development [7].

2.1.3 PATHOGENESIS

The pathogenesis of prostate cancer has not been elucidated fully but genetic changes have been reported in with phenotypic abnormalities leading to carcinoma. The genetic profile of prostate cancer can be classified into two groups: hereditary PCa and sporadic PCa.

Several studies have shown a familial predisposition for PCa. However, identifying causative genes is difficult due to several factors. The advanced age of onset of PCa makes it difficult to study more than two generations. It can be difficult to differentiate sporadic and hereditary cases when they present later in life. Hereditary PCa usually doesn’t have a different clinical presentation to sporadic PCa. Despite these challenges, some epidemiologic studies of PCa have identified high-risk alleles at several genetic loci such as 8p22-23 and 1q24-25 [4].

In sporadic PCa, in the early phases of tumour development, gene profiling studies have suggested a role for tumour suppressor genes such as PTEN, P27, RB1, glutathione S-transferase-π (GSTP1) and transcription factor MYC [8]. A different set of genes have been associated with progression and metastasis such as androgen receptor (AR), TP53 and BCL2 [8]. Of particular interest is the androgen receptor, which reflects the sensitivity of prostate cancer to androgens. Prostate cells with AR receptor mutations may have altered sensitivity to androgens. In particular, the length of CAG repeats in the AR receptor gene varies with ethnicity, with Afro-Caribbean with the shortest and Asians with the longest [9]. This correlates
with the risk profiles, with Afro-Caribbean men having a higher incidence and mortality [2].

From a metabolic point of view, cancer cells rely on anaerobic pathways to convert glucose into ATP, even when there is an abundant oxygen supply. This is known as the Warburg effect [10]. Although glycolysis is less efficient, it produces ATP at a rate 100-times faster than mitochondrial oxidation phosphorylation. This results in the production of lactate in prostate cancer which can be detected by a MRI technique called hyperpolarised MRI. In this imaging technique, a patient is injected with hyperpolarised $^{13}$C-labelled pyruvate which is converted to $^{13}$C-lactate in prostate cancer cells which can be detected using MRI[11].

2.1.4 CLINICAL FEATURES

Many men with prostate cancer have no symptoms at presentation. In some who have large prostates, lower urinary tract symptoms such as hesitancy, dribbling, nocturia, incomplete emptying may exist. Some patients present with haematuria and haematospermia. In patients presenting with metastatic disease, back pain and cachexia might be a feature.

2.1.5 INVESTIGATIONS

The initial assessment of prostate cancer has traditionally been by a serum prostate-specific antigen and digital rectal exam (DRE), which is performed opportunistically in primary care [12]. If these tests are suspicious (increased PSA or nodular prostate), then patients would proceed to a transrectal ultrasound-guided biopsy (TRUS). However, the sensitivity of DRE is low (37%) and there is high inter-observer variability.

The prostate-specific antigen is an enzyme that is produced by prostate epithelium and its role is to make seminal fluid less viscous [12]. A small amount of PSA leaks from the prostate into the bloodstream normally. When the prostate epithelium is
disrupted, more PSA can leak into the bloodstream. Infection, inflammation, benign prostatic hyperplasia, prostate malignancy can cause increased PSA levels [12]. Therefore, PSA is not specific to prostate cancer. Indeed, some cancer types such as neuroendocrine tumours do not secrete PSA and could be missed.

Although PSA is not specific to prostate cancer, it is useful in clinical practice. To account for BPH, a PSA density (PSAD) can be calculated by dividing the PSA level by the gland volume and helps improve the specificity[13]. PSA kinetics are useful in determining recurrence after treatment [14].

Due to PSA being an imperfect biomarker, the evidence for PSA screening is contradictory in the current literature [15,16]. However based on two meta-analyses which showed PSA screening leads to a small reduction in the risk of dying from prostate cancer over 10 years, the US Preventive Services Task Force (USPSTF) and the European Association of Urology support the use of PSA as a screening tool but not in the UK [17,18].

TRUS biopsy samples the prostate under ultrasound guidance but is not targeted based on an area of suspicion. However, due to the transrectal approach, the anterior, peri-urethral, and extreme apex of the prostate are difficult to sample. These limitations result in a considerable false-negative rate, with many low-grade biopsies being upgraded on subsequent prostatectomy[19]. It also carries a small but important risk of sepsis. Localisation of tumours is also sub-optimal due to tangential trajectories.

This triple assessment approach with increased PSA screening led to increased detection of prostate cancer, and it can be argued overdiagnosis of up to 45% of men diagnosed with prostate cancer. Furthermore, the increased detection has also led to overtreatment of low-volume and indolent tumours [20]. A TRUS-guided biopsy also misses clinically significant cancer in up to 30% of cases [21].
2.2 The role of MRI in the assessment of prostate cancer

With the limitations of triple assessment, multiparametric MRI has been used to improve sensitivity and specificity in detecting clinically significant prostate cancer. The first use of MRI in the prostate cancer diagnostic pathway was in the 1980s and involved T1 and T2-weighted sequences. Evidence of MRI in the assessment of prostate cancer has been provided by several studies and meta-analyses published in the last decade. The PROMIS study provided level 1 evidence of diagnostic accuracy of mpMRI before biopsy in men suspected of prostate cancer. Specifically, the study showed that mpMRI had greater sensitivity than TRUS-guided biopsy and a higher negative predictive value for detecting prostate cancer ≥3+4 or cancer core length ≥4 mm [22]. A meta-analysis of 16 studies reported that the use of mpMRI before targeted biopsy resulted in 20% more clinically significant prostate cancer to be identified compared to the TRUS-guided biopsy [23].

Currently, most multiparametric prostate protocols use additional sequences complementing anatomical T2 weighted sequences such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced imaging (DCE). Diffusion-weighted imaging has been shown to increase the sensitivity and specificity of MRI in detecting prostate cancer when added to T2W imaging in a meta-analysis; (pooled sensitivity 0.76 and pooled specificity to 0.82) [24]. Furthermore, the addition of a high b-value image (e.g. b values 2000 s/mm²) can improve sensitivity and specificity up to 88% and 89% respectively [25]. However, the relatively high specificity in the second study was due to the patient population and definition of prostate cancer. These participants in the study were selected to have radical prostatectomy before enrolment and any prostate cancer including Gleason 3+3 was considered positive.

Dynamic contrast-enhanced MRI is performed following injection of a gadolinium-based contrast agent (GBCA) with fat-suppressed T1-weighted acquisitions. The contrast agent allows the characterisation of blood flow to the prostate and
washout. Prostate cancers demonstrate early enhancement due to angiogenesis and early washout. Although inflammation and hyperplastic nodules can mimic this enhancement pattern, the reported sensitivity and specificity for prostate cancer detection is 46-90% and 74-96% [26] respectively.

There have been efforts to standardise the way mpMRI is reported to improve reproducibility and interpretation of the report by Urologists and Oncologists. The initial scoring system was on a 5-point Likert scale, based on a radiologist’s subjective opinion of the likelihood of prostate cancer, which is still used at several UK institutions and is called the ‘Likert Score’ [27]. The ‘Likert Score’ is based on the radiologist’s overall impression based on assessing all the available sequence and clinical information to determine the likelihood of clinically significant prostate cancer and does not specify which sequence is used to determine the score. Prostate cancer has low signal on T2-weighted imaging, high signal on high b value images, low ADC and early post contrast enhancement. Although appearances can vary depending on location and the grade of prostate cancer with higher-grade tumours being more conspicuous and lower grade less conspicuous [27]. This variation in conspicuity can lead to inter-rater variability and therefore a separate scoring system was formulated.

The Prostate Imaging-Reporting and Data system (PI-RADS) was published in 2012 [28]. These criteria defined dominant sequences for the transition zone (T2W) and peripheral zone (DWI). For abnormalities in the peripheral zone, DWI images are the dominant sequence whereas, for lesions in the transition zone, the T2W sequence is the dominant sequence [64,79]. DCE images can upgrade the score for lesions in the PZ if they demonstrate early enhancement. DWI images can upgrade the score for lesions in the TZ. Although this system had reasonable specificity (0.79) and sensitivity (0.78), it was complex and time-consuming. Future iterations in 2014 and 2019 of PI-RADS, called v2 and v2.1, were published in 2014 and 2019 [28,29]. The algorithm was simplified particularly for DCE. A meta-analysis showed that PIRADS v2 performed better than PIRADS v1 in increasing pooled sensitivity (up to 0.95) but not pooled specificity (0.73) [30].
The main differences between the two scoring systems are that the ‘Likert score’ does not specify which sequence is used to score a lesion and radiologists can use clinical information to influence their score (Table 2-1).

This system is particularly useful when dominant sequences are degraded by artefact, but other sequences still raise suspicion for cancer. Both scoring systems have been compared and have shown similar performance or slightly better performance by Likert scoring [31].

Table 2-1 Comparison of Likert and Prostate Imaging Reporting and Data System Scoring

<table>
<thead>
<tr>
<th>Parameter</th>
<th>‘Likert’</th>
<th>‘PIRADS 2.1’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of scoring system</td>
<td>Combination of all sequences, biochemical data and reader experience</td>
<td>Pre-determined imaging features in a specified order</td>
</tr>
<tr>
<td>Scale</td>
<td>5-point</td>
<td>5-point</td>
</tr>
<tr>
<td>Score 1</td>
<td>Clinically significant cancer is highly unlikely to be present</td>
<td></td>
</tr>
<tr>
<td>Score 2</td>
<td>Clinically significant cancer is unlikely to be present</td>
<td></td>
</tr>
<tr>
<td>Score 3</td>
<td>Clinically significant cancer is equivocal</td>
<td></td>
</tr>
<tr>
<td>Score 4</td>
<td>Clinically significant cancer is likely to be present</td>
<td></td>
</tr>
<tr>
<td>Score 5</td>
<td>Clinically significant cancer is highly likely to be present</td>
<td></td>
</tr>
</tbody>
</table>

Both scoring systems can also ascribe an indeterminate score of 3/5 when appearances are not typical for any other score. Many indeterminate lesions
proceed to biopsy. The reported rates of positive biopsy in indeterminate lesions varies from 4% to 33% [32,33]. Some studies have examined combining PSA density and MRI scores to help decide whether indeterminate lesions should be biopsied, with the potential to reduce unnecessary biopsies by 20% [34].

It can be concluded from the body of evidence that the sensitivity and specificity of mpMRI need to be derived from a population who are suspected to have prostate cancer rather than prostatectomy cohorts. In this population such as that studied in PROMIS, mpMRI has high sensitivity and high negative predictive value in men suspected to have prostate cancer but modest specificity.

2.2.1 MPMRI AND BIOPSY

Traditionally biopsy of the prostate was not targeted to a particular lesion but sampled the gland by anatomical regions. However, with the increased use of mpMRI, a newer paradigm of targeted biopsy has been investigated. Several studies have compared the performance of random TRUS biopsy and MRI-targeted biopsy. A randomised controlled study called PRECISION assigned 500 men to either mpMRI or TRUS biopsy and showed that mpMRI-targeted biopsy aided diagnosis of clinically significant prostate cancer in 38% of men, compared with 26% for TRUS-guided biopsy [35]. Not all studies have shown this difference. In two prospective paired-cohort studies, there was no significant difference between these two biopsy techniques [36,37]. A biopsy can be performed via the transrectal or the transperineal approach. Both have reasonable accuracy for targeting lesions, but the trend is towards the transperineal approach due to lower risk of sepsis and better sampling of the anterior prostate [38]. Both biopsy techniques can lead to bleeding post-procedure and 25% of men experience lower urinary tract symptoms after a biopsy [38].

On balance, there is better evidence supporting the use of mpMRI-targeted biopsy than TRUS. This is reflected in the NICE guidelines, which suggest
“multiparametric MRI-influenced prostate biopsy” should be offered to patients with a Likert score greater than 3[39].

2.3 Histopathology

The most common form of prostate cancer is the acinar variant of adenocarcinoma [40]. The other types include transitional cell carcinoma, sarcoma, and neuroendocrine tumours. Most PCa occurs in the peripheral zone (68%) then transition zone (24%) and rarely in the central zone (8%) [40].

The normal prostate microarchitecture displays tubuloalveolar glands which are composed of pseudostratified columnar epithelium surrounding the luminal space. Underlying the superficial epithelial cells are the deeper basal cells which rest on a basement membrane. These cells replace the epithelial cells and regenerate the luminal lining.

This architecture is disrupted in adenocarcinoma. Most adenocarcinomas produce well-defined gland patterns. The abnormal glands are usually smaller than benign glands, but more crowded and lack branching and infolding. Cancerous regions lack the outer basal layer of cells. Like other carcinomas, the nuclei are large and contain one or more large nucleoli. Pre-cursor lesions are referred to as high-grade prostatic intraepithelial neoplasia (PIN). These lesions are characterised by more widely separated, larger branching glands with intra-acinar proliferation but with preserved infolding. These lesions are associated with a higher risk of cancer. They are frequently seen adjacent to or within the same locale as adenocarcinoma. Many of the molecular changes in adenocarcinoma are also seen in PIN lesions [41]. However, the precise clonal and temporal relationship between PIN and adenocarcinoma remains undefined.
There are several grading systems of which the Gleason system is best known, proposed by Dr Donald Gleason\cite{42}. Figure 2-1 shows the original diagram drawn by him.

**Figure 2-1** Original diagram showing Gleason grade and expected morphology by Dr Donald Gleason \cite{42}. 
This system describes five grades based on glandular patterns and degree of differentiation, seen under low magnification. Grade 1 represents well-differentiated tumours with uniform and round glands packed into well-defined nodules. Whereas Grade 5 tumours show no glandular differentiation, and the cancer cells form sheets, nests or cords which infiltrate the stroma. The overall Gleason score is given as a combination of a primary and secondary pattern, for instance, Grade 3+4. This represents the two most prevalent patterns seen in the assessing pathologist and could range from 2 to 10.

Biopsies can be challenging to interpret as there is usually a small volume of tissue containing only a few malignant glands. This is one of the reasons why Gleason patterns 1 and 2 are no longer assigned due to poor reproducibility and poor correlation with radical prostatectomy grade [43]. Most tumours contain more than one pattern and therefore a pathologist can assign a primary grade to the dominant pattern and a secondary grade to the sub-dominant pattern. The two patterns are then added to obtain a combined Gleason score. Combined Gleason scores can then also be put into groups that have similar biological behaviour. The grading groups recommended by the World Health Organisation (WHO), have 5 groups which are Grade Group 1 = Gleason score ≤6, Grade Group 2 = Gleason score 3 + 4 = 7, Grade Group 3 = Gleason score 4 + 3 = 7, Grade Group 4 = Gleason score 8, Grade Group 5 = Gleason scores 9 and 10 [42].

Gleason grading is traditionally a subjective assessment of the specimen by a pathologist. However, with the digitisation of pathology specimens, quantitative assessment is possible. Studies have assessed quantifying the fraction of Gleason 3, 4 and 5 patterns in a specimen or quantifying the percentages of microarchitectural features such as lumen, epithelium, and stroma. One such study showed that the microarchitectural features differ significantly between different Gleason grades [44]. Quantifying the percentage of higher Gleason grades such as Grade 4 and 5 can help whether the patient may be at increased risk of biochemical recurrence.
after prostatectomy \cite{45}. These grade groups have been shown to correlate with 5-year biochemical risk-free survival after radical prostatectomy \cite{46}. Similar prognostic curves were seen in patients treated with radiation ± hormonal therapy.

2.4 MANAGEMENT

The management of prostate cancer depends on whether the disease is localised to the prostate and whether it is a low, intermediate, or high risk according to biopsy results, PSA, and staging.

For high risk localised prostate cancer, patients are offered radical prostatectomy or radical radiotherapy with androgen deprivation therapy. Both have been shown to lower the incidence of disease progression and development of metastases \cite{47}. Metastatic prostate cancer patients are offered chemotherapy and androgen deprivation therapy.

Radical treatment with surgery or radiotherapy can have long term effects on urinary and sexual function. For instance moderate to severe problems with urinary incontinence affect 13% of patients who undergo radical prostatectomy and 5% of patients who undergo radical radiotherapy \cite{48}. Similarly, moderate to severe problems with erectile dysfunction affects 50% of patients who have radical prostatectomy and 36% of patients who have radiotherapy \cite{48}.

Therefore, for low to intermediate-risk localised prostate cancer, there is a trade-off between the morbidity of treatment and the survival benefit. Active surveillance is a recommended option for low-risk disease and comprises regular PSA measurements and interval mpMRI \cite{39}.

Alternative treatments are also available for low to intermediate-risk localised disease, which offers oncological control with the hope of reducing side effects. High intensity focused ultrasound (HIFU), cryotherapy and irreversible
Electroporation are examples of treatment modalities that are used to treat prostate cancer[49,50]

2.5 Novel Imaging Biomarkers (IB) Development and Assessment

Although mpMRI offers well-established advantages compared to the previous diagnostic pathway which consisted of non-targeted TRUS biopsy, there is room for improvement. It is negative in 11% of men with clinically significant cancer, indeterminate in 34% and positive in 30% of men without significant cancer[22]. Overall the specificity of mpMRI has been reported at a modest 37-50%[22,51]. To improve the technique, biomarkers derived from imaging have been developed. The following section explores the framework for biomarker translation in the context of prostate cancer.

The definition of a biomarker has been established as a ‘defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention, including therapeutic interventions’[52]. Although biomarkers such as blood tests are prevalent in modern medicine, few imaging biomarkers are used clinically. Even though many imaging biomarkers have been used as research tools for assessing disease or treatment response, not many have become a ‘clinical decision-making tool’. Examples of the few that have crossed this gap include TNM staging for tumours[52] and PIRADS for prostate cancer[29]. One of the main reasons for this is the lack of rigorous biological and clinical validation of many biomarkers with a prospective clinical trial.

The imaging biomarker roadmap sets out steps that are required to develop an IB with a robust validation[53]. The recommendations suggest a parallel approach of technical validation, biological and clinical validation, and cost-effectiveness analysis. This contrasts with non-imaging biomarkers where clinical validation follows technical validation.
2.5.1 Technical Validation

Technical validation of IBs involves testing precision and bias. Repeatability and reproducibility studies address the precision of an IB. Repeatability studies comprise of repeated measurements in the same subject using the same equipment, processing, and post-processing whereas reproducibility studies use different equipment or different sites in the same or different men.

The scale of these studies depends on the stage of development. For instance, for VERDICT MRI, single centre repeatability of repeating the sequence twice on the same man could be sufficient as a research tool. Single centre reproducibility would comprise of scanning the same man on different scanners. However, for clinical translation, multicentre reproducibility across multiple expert and non-expert centres is required.

Bias is the systematic difference between a measurement and its real value. For many IBs it can be difficult to obtain ground truth values of biological activity from patients without invasive tests. In the context of prostate cancer, some bias can be inferred from histology with histological correlations.

2.5.2 Biological and Clinical Validation

Biological validation is proving the association of the biomarker with a biological variable. For example, an IB increases as the tumour size increases. Clinical utility is different and relates to improvements in health outcomes or in this case, improves diagnostic accuracy.
Biological validation in most cancer imaging uses histology as the gold standard. When this is not available, then pre-clinical tumour models can be used to infer biological association.

In the context of prostate cancer, because biopsy had been the mainstay of diagnosis, many clinical validation studies use histology. Histological outcomes are derived from biopsy or prostatectomy specimens. As explored previously that there is some variability in the biopsy technique, and this can lead to a bias or sampling error. Whole gland histology from prostatectomies provides a more robust reference standard however may not be translatable to the population at large as men undergoing prostatectomy have already been stratified. Therefore, for diagnostic studies, standardised biopsy in a biopsy naïve population supplemented by whole gland histology in those men who opt for surgery is a better reference standard.

Clinical utility usually happens late in the development of an IB after multicentre technical validation has been performed. Clinical utility in the context of prostate cancer diagnosis would address unmet clinical needs in the current pathway. For instance, before mpMRI was introduced, there was overdiagnosis of insignificant prostate cancer leading to overtreatment and underdiagnosis of significant cancer leading to poor patient outcomes by blind transrectal biopsy. The introduction of mpMRI in the diagnostic pathway addressed both needs as shown by prospective clinical trials [22].

Currently, the two clinical needs that IBs could address in prostate cancer are:

1) Avoid unnecessary biopsies in men who have benign or clinically insignificant cancer
2) Avoid unnecessary treatment in men who have indolent cancer

2.5.3 COST-EFFECTIVENESS
For biomarkers to be translated, there needs to be an assessment of cost-effectiveness. Imaging especially MRI can be an expensive test and if it does not add more clinical utility compared to less expensive serum or biofluid biomarkers, it may not be translated. For IBs, it can be difficult to show an advantage in terms of Quality-adjusted life year (QALY). However, in prostate cancer assessment, if an IB can help avoid unnecessary biopsy, this can have an impact on QALY.

2.5.4 VERDICT BIOMARKER DEVELOPMENT

The current stage of development of VERDICT MRI as a quantitative imaging biomarker is shown in Figure 2-2.
Figure 2-2 Current stage of biomarker development for VERDICT MRI

The boxes highlighted in green are finished studies. The boxes highlighted in yellow are addressed in this thesis. The orange boxes are ongoing. The box in red has not been addressed.

2.5.4.1 Biomarker Validation

Initially, VERDICT MRI was studied in tumour xenograft models of colorectal cancer. This study showed that VERDICT MRI could identify significant changes in cell size and vasculature which occurred after the administration of a chemotherapeutic agent, in contrast to standard ADC and IVIM models [54].
VERDICT MRI was then investigated in 8 patients with biopsy positive prostate cancer.

This feasibility study showed that VERDICT could distinguish between benign and cancerous regions with differences in VERDICT parameters [55]. A further study in five patients undergoing prostatectomy, related to VERDICT derived parameters from ex-vivo prostate MRI to histopathological appearances [56]. This study found that VERDICT derived intracellular fraction corresponded to histopathological indicators of cellular fraction and the collagen fibre patterns in stroma correlated with the tensor analysis from the ex-vivo scans. VERDICT has also been investigated in other tumour types such as gliomas and bone metastases [57,58].

2.5.4.2 VERDICT Technical Validation

A repeatability study was performed on a subset of patients (n=42) as part of the INNOVATE study [59]. Most participants (n=31) underwent a back-to-back scan whereas 10 patients were scanned with a 5-minute gap between scans. In this interval, participants walked around the scanner room. The repeatability of various VERDICT parameters was examined using the intraclass correlation coefficient. FIC (ICC 0.89-0.92) and FEES (ICC 0.86-0.880 from VERDICT had ‘almost perfect’ repeatability whereas FVASC (ICC 0.81-0.83) had the lowest repeatability. Reproducibility of VERDICT MRI remains to be performed.

2.5.4.3 Cost-effectiveness

This analysis has not yet been performed because of the early developmental stage of VERDICT MRI. Cost-effectiveness depends on how VERDICT MRI may be applied in a clinical setting, which has not been determined yet. The work presented in this thesis assesses the possible ways VERDICT MRI could be
implemented. Future work will include a cost-effectiveness analysis based on the recommendations from this work.
3 Introduction to Magnetic Resonance Imaging

The following books were reviewed and used in this section:


3.1 The Nuclear Magnetic Resonance Signal

The source of the MRI signal is based on nuclear magnetic resonance. It can be explained by both ‘classical’ physics and quantum mechanics. Atomic nuclei have an intrinsic quantum property called spin. The ‘classic’ description of spin is a particle spinning on its axis with angular momentum.

To exhibit nuclear magnetic resonance, a nucleus needs to have a non-zero spin quantum number (an odd number of protons/neutrons). For the hydrogen nucleus which is a singular proton and has a non-zero quantum number, the two directions or values that the angular momentum can be projected in is either -½ or ½, either aligned with the external field $B_0$ or aligned in the opposite direction. These are designated as ‘spin down’ for protons antiparallel to the field and ‘spin up’ for protons aligned to the external field. Transitions between these energy levels can occur if an appropriate packet or quanta of energy is applied. The energy in the form of an electromagnetic wave needs to be of a particular frequency, governed by the following equation:

$$f = \frac{\gamma B_0}{2\pi} \text{ or } \omega = \gamma B_0$$
Where \( f \) is the frequency of the electromagnetic wave in Hertz, \( \gamma \) is the gyromagnetic ratio and \( B_0 \) is the strength of the external magnetic field. The frequency can also be expressed as an angular frequency \( (\omega) \) in radians per second. This expression is known as the Larmor equation.

In the classical description, when a nucleus that has a magnetic moment (due to its positive charge) is placed in an external magnetic field, it experiences a torque that causes it to precess. The angular frequency of the precession is also described by the same Larmor equation.

3.1.1 **NET MAGNETISATION**

For a large number of protons such as within the human body, the ‘classical’ description is useful. In the absence of an external field, the vector sum of all the magnetic moments from protons is zero as the randomly aligned moments cancel out. However, when an external field is applied, protons are either in ‘spin up’ or ‘spin down’, with more in spin up or the lower energy state. This results in a net magnetic moment that is parallel to the external field \( (B_0) \). The distribution is described by the Boltzmann distribution when in thermal equilibrium. For instance, at body temperature \( (37.4^\circ\text{C}) \), there is one extra spin-up proton out of a million protons at 1.5 T. In a human body, this difference when summated over the entire proton population gives the net-magnetisation-vector (NMV).

3.1.2 **EXCITATION**

The net magnetisation vector is parallel to the external magnetic field and therefore difficult to detect as it’s of a much smaller magnitude. To measure the net magnetisation, this needs to be projected or ‘tipped’ in the perpendicular/transverse plane to be detected as a signal. The required frequency to cause the net
magnetisation to ‘tip’ into the transverse plane in the classical description is in the order of radiofrequency waves in the electromagnetic spectrum.

The movement of the magnetic moment which is already precessing can be simplified in a rotating reference frame which itself rotates in the axis of the external field. This allows the angular rotation of the magnetic moment to appear stationary and simplifies the description when it is acted upon by radiofrequency radiation. The non-rotating frame is called the laboratory reference frame with axes x, y and z whereas the rotating reference frame has axes x’, y’ and z’. By convention, the external magnetic field $B_0$ is along the z’ axis and the rotating frame also rotates around the z-axis.

The motion of the net magnetisation due to applied magnetic fields (radiofrequency pulse), is described by the Bloch equation:

$$\frac{dM}{dt} = M \times \gamma B$$

where M is the magnetisation vector, $\gamma$ is the gyromagnetic ratio and B is the applied field.

The radiofrequency pulse has an oscillating magnetic field perpendicular to the static magnetic field and causes a torque that rotates the net magnetisation towards the transverse plane. The angle of rotation is called the flip angle.

If the radiofrequency pulse is applied for the appropriate time, it results in the net magnetisation vector rotating fully in the transverse plane and is called a 90° pulse. The transverse component called $M_{xy}$ is equal in magnitude to the initial $M_z$. This rotating magnetic moment induces a current in a radiofrequency coil by magnetic induction and we can detect the NMR signal.
3.1.3 RELAXATION

The spin system returns to equilibrium after the energy is applied and this can be thought of as having two components, namely longitudinal and transverse relaxation.

3.1.3.1 Longitudinal Spin-Lattice relaxation

After excitation, there are protons in a higher energy level that are aligned against the external magnetic field. These protons then transition to a lower energy level and emit a photon or quanta of energy. This process is called Stimulated Emission. This emitted energy is absorbed by the surrounding milieu, called the lattice. The time constant for this is defined as $T_1$, described by the following equation:

$$M_z(t) = M_0(1 - e^{-\frac{t}{T_1}})$$

where $M_0$ is the initial net magnetisation, $T_1$ is a time constant, and 't' is the time after the 90° pulse (Figure 3-1).
From a classical point of view, $T_1$ describes the regrowth of the net magnetisation ($M_z$) along the ‘z’ axis as spins realign with the static magnetic field $B_0$. As this depends on the surrounding lattice, $T_1$ values differ depending on tissue type.

### 3.1.3.2 Transverse Spin-spin $T_2$ relaxation

Transverse relaxation is caused by the exchange of energy between adjacent protons/spins. Each spin acts as a small magnet and has a local magnetic field which can alter the spin precession frequency of other spins. These spin-spin interactions cause *dephasing*. This is the process where the spins progressively lose sync with each other and rotate at different speeds due to experiencing local fields. Spin-spin relaxation is also affected by molecular motion or tumbling. For instance,
in liquids where molecules move more freely than in solids, the local magnetic field experienced by a particular molecule is more homogenous due to averaging of local fields. In contrast, molecules in solids are more fixed and experience local field differences leading to rapid dephasing. The time constant $T_2$ describes the decay of $M_{xy}$ to 0 and is longer for liquids and short for solids. It is described by the following equation:

$$M_{xy}(t) = M_0 e^{-\frac{t}{T_2}}$$

where $M_{xy}$ is transverse relaxation, $T_2$ is the time constant, and $t$ is the time following the 90° RF pulse (Figure 3-2). The $T_2$ value varies between different types of tissue, but also between pathological states of the same tissue. The $T_2$ of prostate cancer in the peripheral zone is lower compared to the normal peripheral zone.

The magnetic field is not perfectly homogenous, and some tissues have paramagnetic properties which cause faster dephasing. Each spin, therefore, experiences a different magnetic field depending on position and environment. This decay is known as the Free Induction decay (FID). The time constant for this is $T_2^\ast$.

### 3.1.3.3 Chemical Shift

The electrons surrounding the nucleus also has an impact on the magnetic field experienced by the nucleus. The electron cloud around nuclei results in a small magnetic field that opposes the direction of the external field. This causes a shielding effect and reduces the field experienced by the nucleus. This shielding effect depends on the distribution of electrons, surrounding atoms and atomic bonds. This is demonstrated in a slight difference between the precession frequency of protons in fat versus water. The protons in a triglyceride for example is surrounded by a greater density of electrons than in water.
3.2 Pulse Sequences

The previous section explained the fundamentals behind an NMR signal, the next section discusses how an image can be formed from acquiring NMR signals from protons.

3.2.1 Free Induction Decay

Immediately after excitation by a 90° RF pulse, a transient oscillation can be detected in a receiver coil due to transverse tipping of the NMV. This signal is known as the Free Induction Decay (FID). The frequency of oscillation is at the Larmor frequency and the time constant of the exponential signal is $T_2^*$, as shown below (Figure 3-3). The FID can be analysed by Fourier Transformation (FT) which gives a frequency component of the signal. The FID usually does not contribute to the MR image but is useful for MR spectroscopy. The signal produced by gradient or spin-echo sequences is an ‘echo’ which is produced at a pre-determined interval after the initial excitation. This allows time for spatial localisation to take place. The mechanism of these two sequences is discussed later.
The analogue signal acquired at the time to echo is then digitised via a method called Discrete Fourier transformation. The analogue signal is sampled at regular intervals to accurately capture the information. The frequency of sampling needs to be high enough to preserve the fidelity of the signal and cover the frequency range within the signal. The sampling frequency can be determined by the Nyquist Theorem. This states that to accurately sample a signal which contains no frequencies higher than ‘F’ Hz, it can be completely determined by giving its magnitude at a series of sampling points spaced 1/(2F) seconds apart.

3.2.2 SPATIAL LOCALISATION

3.2.2.1 Slice Selection
As in other cross-sectional imaging, a volume of tissue is imaged by repeated slices of a set width which can then be reconstructed into a 3-D volume. To excite a slice of tissue in a subject, a ‘slice select gradient’ is activated in conjunction with a tailor-made RF pulse with a limited range of frequency components. The slice select gradient linearly varies the potential resonant frequencies of the spins in the slice of interest which matches the frequency profile of the tailored RF pulse, allowing selective excitation. The thickness of the selected slice depends on the gradient amplitude of the slice select gradient and the bandwidth of the RF pulse. To reduce slice thickness, the gradient amplitude can be increased, or the bandwidth is reduced. The former is the preferred method in most systems for 2-D multi-slice imaging.

3.2.2.2 Frequency Encoding

Slice selection provides spatial information in one axis and further gradients are required to obtain information from the other axes. By applying a ‘frequency encoding gradient’ in a given axis, each spatial location along the gradient direction has a unique resonance frequency. When the spatial frequency in the object matches that of the applied gradient, a signal is detected. The signal is collected whilst the frequency encoding gradient is activated which is why it is also called a ‘read-out’ gradient. The acquired signal is a collection of spatial frequencies in a given amount of time which are then separated by a Fourier transform.

3.2.2.3 Phase Encoding

By applying an additional gradient before and perpendicular to the frequency encoding gradient, spatial information can be obtained in another dimension without interfering with frequency encoding. This causes a phase change that varies linearly with position and persists after the gradient is deactivated. This gives spatial information in a dimension perpendicular to the frequency encoding gradient. This step is usually repeated several times with the imaging experiment to encode the entire field of view.
3.2.2.4 k-space

The information from the spatial encoding of the imaging experiment is stored in a data matrix known as k-space. K-space is a representation of spatial frequencies which have been derived from the readout MR signal. This matrix is filled row by row by the read-out or frequency encoding gradient for each phase encoding step, in conventional Cartesian acquisition. The axes of k-space are $k_x$ and $k_y$, representing the frequency and phase encoding axes. In the centre of the k-space, there are lower spatial frequencies, and the edge has higher spatial frequencies. To generate an image from this collection of spatial frequencies, inverse two-dimensional Fourier transformation is performed.

3.2.2.5 The Echo

Gradient coils have a ramp up time which is the time taken for the gradient to be induced. Immediately after the 90° RF pulse, the gradient is changing in size and if the signal is sampled at this point, the spatial localisation would be inaccurate. Imaging sequences are designed so that an ‘echo’ of the NMR signal occurs at a later point to allow for time for the spatial localisation gradients to activate and become stable. Two different types of echo forming sequences are now discussed.

3.2.2.6 Gradient Echo

In a gradient echo pulse sequence, gradients are used to de-phase and then rephase spins to create an ‘echo’ at a predictable time after excitation. Initially, a negative gradient lobe is applied immediately after excitation which causes a dephasing effect. This is followed by a positive gradient which is applied for twice the duration of the negative gradient. The positive gradient causes rephasing and in the middle of the positive gradient, the spins come back into phase and form a coherent maximum signal or echo.
In the time after the initial 90° pulse, relaxation of the signal is also occurring. In particular, T2* decay which is caused by field in-homogeneities and spin-spin relaxation causes a reduction in the magnitude of the signal. This is described by the equation below. Therefore, gradient-echo images have a T2* weighting.

\[ M_{xy}(t) = M_0 e^{-\frac{TE}{T_2^*}} \]

3.2.2.7 Spin Echo

Although T2* weighting can be useful, true T2 weighting needs to overcome the rapid dephasing effects of field inhomogeneities which contribute to T2* decay. A 180° refocusing RF pulse can remove these dephasing effects, by ‘flipping’ the spins by 180°and reversing the phase angles of the spins in the transverse plane. The effect of this is that the spins gain coherence and form an echo at time TE.

\[ M_{xy}(t) = M_0 e^{-\frac{TE}{T_2}} \]

The signal amplitude at TE, therefore, is given by the above equation and is based on T2 decay. Several 180° refocusing pulses can be applied until M_{xy} reaches 0 due to T2 decay, before re-applying the 90° excitation RF pulse. Clinically, this sequence is used to image the prostate due to higher signal to noise and favourable tissue contrast compared to gradient-echo sequences.

3.2.2.8 Image Contrast

Image contrast depends on the amplitude of the acquired MR signal at time TE (time to echo). It depends on proton density, T1, T2 or T2* relaxation. For a spin-echo sequence, it can be expressed as:
\[ S = PD \left[ 1 - e^{-\frac{TR}{T_1}} \right] e^{-\frac{TE}{T_2}} \]

Where S is the signal intensity, PD is proton density, TR is time to the repetition of the 90° RF pulse and TE is time to echo.

By changing TR and TE values, we can determine which component of the equation dominates. For PD-weighted imaging, a long TR compared to \( T_1 \) and a short TE compared to \( T_2 \), minimises the effect of \( T_1 \) and \( T_2 \). For \( T_2 \) weighting, we minimise \( T_1 \) effects by having a long TR and choosing a moderately long TE. Finally, for \( T_1 \) weighting, we can choose a short TE and a short TR.

### 3.2.2.9 Fat Suppression

A large proportion of protons in the human body reside in fat and water molecules. Sometimes it is useful to emphasise these components to detect pathology. In prostate imaging, fat suppression is useful in dynamic contrast-enhanced imaging. Both fat and contrast enhancement can cause hyperintense signal and so suppressing fat leads to better contrast for detecting enhancement.

### 3.2.2.10 Inversion Recovery

Inversion recovery such as that used in short tau inversion recovery (STIR) imaging involves an inversion pulse before signal readout. This has the effect of rotating the magnetisation vector by 180°. The recovery of \( M_z \) starts from \(-M_0\) and depends on \( T_1 \) relaxation. Protons in fat recover faster compared to protons in other tissue and therefore can be separated before the 90° pulse. The timing of the 90° RF pulse is chosen to coincide with the time the fat signal reaches 0. This time is called time to inversion or TI (Figure 3-4). This process is called *selective nulling*. 
The signal at the time to echo is from the non-fat protons in the tissue being examined.

### 3.2.2.11 Chemical shift-selective excitation (CHESS)

This strategy uses selective excitation of the fat signal by a tailored RF pulse, followed by a *crusher* gradient to negate the resultant transverse magnetisation. The rest of the sequence follows and the protons in fat do not contribute to the overall MR signal.

![Inversion Recovery for fat suppression](image)

**Figure 3-3** Inversion Recovery for fat suppression

### 3.2.2.12 SPIR and SPAIR

Spectral Pre-saturation with Inversion Recovery (SPIR) uses a combination of the above-described techniques for fat suppression. It utilises an inversion RF pulse with a flip angle of 100 to 180°, and a spoiler gradient to negate any transverse magnetisation from protons in fat. The Spectrally Attenuated Inversion Recovery (SPAIR) uses a different inversion pulse called an adiabatic pulse which is
frequency and amplitude modulated and less sensitive to field inhomogeneities. These techniques are commonly used to suppress the fat signal in diffusion-weighted imaging of the prostate.

3.2.3 DIFFUSION WEIGHTED IMAGING

3.2.3.1 Brownian Motion

Diffusion is the random displacement of molecules in a fluid, under the influence of temperature, viscosity, particle size and the presence of a physical barrier. The movement of molecules that displace and collide with others is seen as a ‘random walk’ described by Einstein. This displacement when studied at a population level is seen to have a Gaussian distribution. Einstein showed that the mean square displacement (r²) is given by:

\[ < r^2 > = 6Dt \]

where D is the diffusion coefficient, and t is the observation time. These distances are in the order of microns which is a very useful order of magnitude to interrogate tissue.

The diffusion coefficient can be calculated by the Stokes-Einstein equation:

\[ D = \frac{k_BT}{6\pi\eta r} \]

where T is absolute temperature, \( k_B \) is the Boltzmann constant, \( \eta \) is dynamic viscosity and r is the radius of the spherical particle.
At a cellular level, the diffusion environment is different according to the compartment a water molecule is in. In the intracellular compartment, the cell membrane, intracellular proteins, and other macromolecules have an influence and cause restriction of diffusion. In the extracellular space, vessels and cell membranes can restrict diffusion. In certain pathological states, these relationships change which is reflected in differences in diffusivity. This makes diffusion a useful tool in evaluating disease.

3.2.3.2 Diffusion Sensitive MRI

To sensitise a spin-echo sequence to diffusion, two identical diffusion-sensitising gradients are added on either side of the 180° refocusing pulse (Figure 3–5). The effect of these identical gradients is to change the phase of the spins. If the spins are stationary, the phase change acquired during the first gradient is cancelled out by the second gradient. In contrast for moving spins, the phase gained during the first gradient is not equal to that lost during the second gradient pulse. At the time to echo, this phase difference causes dephasing, and the magnitude of the signal is reduced.

\[ \delta = \text{time diffusion gradient is on}, \Delta = \text{time between diffusion gradients}. \ G_{\text{diff}} = \text{Diffusion gradient} \]
3.2.3.3 Echo Planar Imaging

Modern DWI acquisitions use a fast-imaging method called echo-planar imaging. Fast imaging in the context of diffusion is useful as it reduces macroscopic motion artefacts.

One of the main reasons why conventional spin and gradient echo sequences take time is because each phase encoding step requires the repetition of the pulse sequence. In echo-planar imaging, the same phase encoding information can be acquired in a ‘single shot’ or after a single RF excitation pulse (Figure 3-6). This is done by first applying a large negative or positive phase encoding gradient, followed by smaller phase encoding gradients, also called blips. These smaller phase encoding gradients have the effect of adding or subtracting from the initial larger phase encoding gradient, typically allowing 128 phase encoding steps to be completed. In addition, after the initial excitation pulse, stronger oscillating frequency-encoding/read-out gradients are applied which allow gradient echoes to be collected with each oscillation. Therefore k-space is filled in a ‘zig-zag’ trajectory rather than row by row seen in conventional cartesian trajectories (Figure 3-7).

Figure 3-5 Echo Planar Pulse Sequence
The spin-echo signal magnitude is given by:

\[ S = S_0 \cdot e^{-bD}, \]

where \( S_0 \) is signal magnitude before applying diffusion gradients, \( b \) is diffusion weighting and \( D \) is the diffusion coefficient. Diffusion weighting or ‘b’ is given by:

\[ b = \gamma^2 G_{diff}^2 \delta^2 (\Delta - \frac{\delta}{3}) \]

where \( \gamma \) is the gyromagnetic ratio, \( G_{diff} \) is the strength of the diffusion gradient, \( \delta \) is diffusion gradient duration and \( \Delta \) is the diffusion gradient separation.

As the diffusion sensitising gradients increase in strength/duration, the signal intensity decreases as protons in water diffuse. In an object with free water diffusion, the decrease is mono-exponential as shown below (Figure 3-8). However,
in a more complex environment such as the prostate, water molecules are in different compartments and environments with different diffusivity. The decay of the signal in tissue is multiexponential. Several models have been proposed to describe this decay.

### 3.2.3.4.1 Apparent Diffusion Coefficient

The simplest model assumes a mono-exponential decay of the signal. The diffusion coefficient can be calculated using the formula below. The calculation requires acquiring diffusion-weighted images at two or more different b-values. The ADC can then be calculated by solving for ‘D’/‘ADC’. For instance, if $b=1000$ and $b=0$ images are acquired, the ADC is given by:

$$ADC = \frac{\ln (S_{b=1000}/S_{b=0})}{1000}$$

This is a useful quantitative measure because the diffusion pulse sequence is sensitive to both diffusion and $T_2$ decay. The ADC can be calculated for each voxel and presented as an ADC map. This image is useful as if a region of high signal on diffusion-weighted images corresponds to a hypointense or low ADC, then this represents true restricted diffusion, rather than high $T_2$ values from the tissue ($T_2$ shine through). The reason why the diffusion coefficient is referred to as ‘apparent’ when measured this way is that in tissues there are multiple compartments that can affect the ADC and not just the diffusivity of water. More complex models for diffusion are now explored.
3.2.3.4.2 Intravoxel Incoherent Motion (IVIM)

A component of the signal decay is thought to be due to perfusion. Although blood microcirculation and diffusion are quite different in spatial and temporal scales, there are similarities in the bulk movement of blood in the seemingly random network of capillaries. This collective movement of molecules in blood in the microcirculation has been proposed to represent pseudo-diffusion. The values derived for the pseudo-diffusion coefficient are close to the diffusion coefficient and therefore diffusion MRI is sensitive to both perfusion and diffusion. These are separated by the IVIM model and modelled as a biexponential decay:

$$S = S_0[(1 - f)e^{-bD} + fe^{-bD*}]$$

**Figure 3-7** Signal decay curve showing the deviation from the gaussian mono-exponential assumed decay of the signal.

The non-gaussian components and their respective models are highlighted in green and blue lines.
Where $S_0$ is the signal intensity at $b=0$ ss/mm$^2$, $f$ is the flowing blood fraction, $D$ is the water diffusion coefficient in tissue, $D^*$ is the pseudo-diffusion co-efficient\,[60]. The fraction of water molecules in flowing blood is small compared to whole tissue water content, and the so this effect is seen at low $b$ values.

### 3.2.3.4.3 Stretched Exponential Model

The signal decay at higher $b$-values appears to have a different decay profile compared to lower $b$-values. Two main signal models are prevalent in the literature for fitting high $b$-value data called the Stretched Exponential and Diffusion Kurtosis model.

The stretched exponential introduces an additional component called the anomalous exponent $\alpha$, to explain the deviation from the mono-exponential behaviour,

$$S = S_0 e^{-(bADC_s)^\alpha}$$

where $ADC_s$ is the diffusion coefficient of the stretched exponential model, $\alpha$ is the anomalous exponent\,[61]. The anomalous component can vary from 0 to 1; with higher values suggesting high homogeneity in apparent diffusion and therefore close to mono-exponential decay of free water. A lower value reflects heterogeneity and is thought to reflect multiple separable proton pools with a voxel.

### 3.2.3.4.4 Diffusion Kurtosis Model
The deviation of the monoexponential decay curve at high b-values can be modelled by the addition of two variables $D_{\text{app}}$ and $K_{\text{app}}$. $D_{\text{app}}$ represents the corrected diffusion coefficient considering non-gaussian behaviour and $K_{\text{app}}$ is the apparent diffusion kurtosis. A greater kurtosis reflects a more peaked distribution compared to a normal distribution.

### 3.2.3.5 Diffusion Anisotropy and Tensor Imaging

When tissue microstructure causes a preferential direction of water diffusion, measuring the direction of diffusion can give the geometry of the microstructure. This process is called anisotropic diffusion.

The diffusion tensor is a 3 x 3 matrix which contains information about the diffusion co-efficient in 3 orthogonal directions.

$$DT = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}$$

As this matrix contains some elements which are equivalent such as $D_{xy} = D_{yx}$, there are 6 unknown parameters. Therefore at least 6 different diffusion gradient directions are needed to complete the tensor. From a diffusion tensor for each voxel, one can calculate the principal diffusion direction.

### 3.2.4 Microstructural Models and Verdict

Microstructural models use mathematical representations of the microarchitecture of tissue and relate them to the signal obtained from diffusion MRI. These exist for
various tissue types particularly in the brain, but the focus here is the prostate [62].

3.2.4.1 VERDICT

Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumours is a microstructural model which has three main compartments [54]. These compartments describe the diffusion signal in three separate populations of water molecules from (1) water trapped in cells ($S_1$), (2) water in the interstitium ($S_2$) and (3) water in vessels ($S_3$).

Mathematically, these compartments represent three parametric models which are summated with VERDICT. The signal for the multi-parametric VERDICT model is given by:

$$S = \sum_{i=1}^{3} f_i S_i$$

where $f_i$ is the proportion of signal with no diffusion weighting ($b=0$) from water molecules in population $i, 0 \leq f_i \leq 1, \sum_{i=1}^{3} f_i = 1$.

3.2.4.2 Intracellular compartment

Diffusion in the intracellular compartment is restricted by cell membranes, internal boundaries such as nuclei and organelles, macromolecules, and other structures. The diffusion here can be considered isotropic as there is no direction diffusion is
restricted in unlike in the vascular compartment. Mathematically, the VERDICT model uses spheres that have impermeable boundaries. The parameters in this compartment are $f_{IC}$ (intracellular volume fraction), $d_{IC}$ (intracellular diffusivity), and cell radius $R$.

### 3.2.4.3 Interstitium

The interstitium also called the extra-cellular and extra-vascular space (EES), is based on the diffusion tensor model and free Gaussian diffusion. The water in luminal spaces of the prostate is relatively free, however, water molecules in tissue fluid in the stroma and extracellular matrix are likely to be non-gaussian. Mathematically, this is modelled by isotropic tensor ‘balls’ with Gaussian diffusion. The parameters in this compartment are $FEES$ (EES volume fraction) and $d_{EES}$ (diffusivity EES).

### 3.2.4.4 Vascular space

The diffusion in individual vessels is the faster pseudo-diffusion and usually in a distinct direction. However, in a capillary network, a single voxel may contain more than one vessel with different directions of flow. This is modelled mathematically, by an ‘astrostick’ which is a cylinder with anisotropic pseudo-diffusion, but the overall sum in each voxel is considered to be isotropic. The parameters in this compartment are $FVASC$ (vascular volume fraction) and $P$ (pseudo-diffusivity).

By fitting this model to the signal from each voxel from a variety of $b$ values and diffusion timings, several parameters can be computed. Parameter maps are also derived to represent the three compartments such as $f_{IC}$, $FEES$ and $FVASC$. 

65
These compartments are thought to be significantly perturbed by the presence of prostate carcinoma. The glandular density increases with the proliferation of epithelial cells which increases the intracellular compartment. There is neovascularisation which should impact the vascular compartment. The stroma also changes in cancer with decreased luminal space and increase in vessels which could affect the extravascular extracellular space.
4 Clinical outcomes from the INNOVATE study

Author declaration

All the work in this chapter was conceived and written by me, under the supervision of Professor Shonit Punwani. Ethical approval was obtained by Dr Edward Johnston and Professor Shonit Punwani. Patient recruitment was performed by myself, Joey Clemente and Dr Edward Johnston. The clinical data were collected by myself and cross-checked with Dr Hayley Pye who independently collected clinical data for a different component of the trial involving serum and fluidic markers. Multiparametric MRI was read by a group of 10 uro-radiologists. Biopsies were performed by urologists at UCLH and Barts. Histopathology was reported by consultants at UCLH and Barts Health.

4.1 INTRODUCTION

As discussed in earlier chapters, multiparametric MRI has been a recent introduction in the UK to the diagnostic pathway and was recommended as a first-line investigation for all men with a suspicion of clinically localised prostate cancer by NICE in 2019. This was based on studies that showed that mpMRI in biopsy naïve patients had greater sensitivity and higher negative predictive value for detecting clinically significant prostate compared to the previous pathway which used TRUS guided biopsy and prostate-specific antigen (PSA) [22]. However, as mentioned before there is room for improvement. It is negative in 11% of men with clinically significant cancer, indeterminate in 34% and positive in 30% of men without significant cancer [22]. Due to this, a few national bodies such as American Urological Association and Cancer Care Ontario have not included it in their national guidelines.

One of the reasons for low specificity is that the sequences which make up multiparametric MRI are limited in their differentiation between benign conditions and cancer. For instance, prostatic atrophy and inflammation can cause abnormal signal in all sequences and lead to false positives [69]. Biopsies can examine the
microstructure of the prostate and therefore distinguish between benign and malignant. Diffusion-weighted imaging can also provide information about the microstructure of the prostate. In its current form, a high b-value image and ADC maps are used to detect restricted diffusion in the prostate. ADC is a measure of the magnitude of diffusion of water within tissue and is low in tumours. However, it can also be low in benign conditions such as atrophy and inflammation [63]. Inflammation can also mimic the early enhancement seen on dynamic contrast-enhanced imaging. These conditions lead to false-positive mpMRI results and unnecessary biopsies.

VERDICT MRI as discussed earlier is a more complex mathematical model of the acquired diffusion signal. The model provides quantitative biophysical parameters which are akin to histological features and could be more specific than ADC. To investigate early clinical utility, a prospective cohort study was designed called INNOVATE, Combining advances in imaging with biomarkers for improved diagnosis of Aggressive prostate cancer.

The INNOVATE study is a prospective, cohort study evaluating the value of additional VERDICT MRI to the mpMRI-directed diagnostic pathway for suspected prostate cancer [64]. The study gained ethical approval in December 2015 (London-Surrey Borders 15/LO/0692) and recruitment started in April 2016. The study design also included the collection of blood and urine (fluidic markers) from participants. The design of the study allowed patients to follow the standard of care diagnostic pathway, with the addition of VERDICT MRI and fluidic markers before any biopsy is carried out. Results from VERDICT MRI or fluidic markers were not used to make biopsy decisions.

The standard diagnostic pathway at University College London comprises initial stratification with a clinical serum PSA test and mp-MRI including dynamic contrast-enhanced imaging. The PSA level, mp-MRI Likert score and PSA-density (PSAD) are used to decide whether to proceed to biopsy or not. A biopsy is targeted to specific regions of the prostate based on the MRI findings. If there is no
target, then a template mapping approach is used to sample all the regions of the prostate.

When the mp-MRI is scored as indeterminate (Likert 3/5), the PSAD is used to determine whether a patient is recommended a biopsy. PSAD calculated by dividing the PSA by the mpMRI derived prostate volume. Different PSAD thresholds have been proposed in the literature. The most widely used is 0.15 \[65,66\] based on retrospective analysis, however, this has not been validated prospectively. A lower threshold has also been proposed at 0.08 from a different retrospective analysis of 2162 men \[67\]. At UCLH, PSAD is used in counselling the patient but not used stringently.

This chapter will present the clinical diagnostic outcomes, in particular, PSAD, multiparametric MRI stratification and biopsy outcomes of the mpMRI-directed pathway and set a baseline to which comparisons can be made for VERDICT MRI. Both Likert scoring and PIRADS 2.1 scoring will be presented.

4.1.1 AIMS

i) Test PSA in differentiating clinically significant cancer

ii) Test PSAD in differentiating clinically significant cancer

iii) Test Multiparametric MRI in differentiating clinically significant cancer

4.1.2 NULL HYPOTHESES

i) The distribution of PSA is the same across clinically significant cancer and benign/non-clinically significant cancer

ii) The distribution of PSAD is the same for patients with clinically significant cancer and benign/non-clinically significant cancer
Multiparametric MRI Likert scores are the same for significant cancer and benign/non-clinically significant cancer

4.2 METHODS

4.2.1 STUDY DESIGN

Full ethical approval for the study was granted by the London–Surrey Borders Research Ethics Committee (15/LO/0692). The INNOVATE study recruited participants from two sites: Barts Health and University College London Hospital Trust. Recruited participants underwent the standard of care diagnostic pathway for suspected prostate cancer. Men were usually referred from primary care due to raised prostate-specific antigen levels or suspicious digital rectal exam.

Men recruited to the study underwent both mpMRI and VERDICT MRI in one scanning session or two separate sessions, but before any biopsy was performed. The VERDICT MRI was not used to make biopsy decisions.

Clinically significant prostate cancer is defined as overall Gleason $\geq 3+4$ of any length. These criteria have been used in multiple studies and reflect the differing management of patients[23,35,68]. Patients who have 3+3 disease are usually put on an active surveillance programme whereas patients with $\geq 3+4$ disease or higher undergo active treatment with prostatectomy, focal therapy, or radiotherapy.

4.2.2 STUDY POPULATION

Study participants were identified between April 2016 to August 2019. Initially, this included men undergoing active surveillance of prostate cancer ($n=42$) which will be used as a training set (Figure 4-1).
For this study, I included men who were biopsy naïve. I excluded a) men unable to have an MRI scan or in whom artefact would reduce quality of MRI (bilateral hip prostheses) b) men unable to give informed consent c) men with previous treatment for prostate cancer (prostatectomy, radiotherapy, brachytherapy b) Men on hormonal treatment for prostate cancer c) Previous prostate biopsy within 6 months of scheduled mpMRI. Men who had images inadequate for analysis due to artefact or faults with image acquisition were withdrawn.

Figure 4-1 Study patient selection

4.2.3 Multiparametric MRI protocol
All participants underwent mpMRI with either a 1.5T or 3.0T scanner with a pelvic-phased array coil as part of standard clinical care. A standard dose of a spasmolytic agent (Buscopan, Boehringer Ingelheim, Ingelheim am Rhein, Germany) was given intravenously before imaging to reduce bowel peristalsis. The mpMRI sequences included T1-weighted (T1W), T2-weighted (T2W), dynamic contrast enhancement (DCE) with gadolinium, diffusion-weighted imaging (DWI, values: 0, 150, 500, 1000 and either 1400 or 2000 s/mm²) and vendor produced apparent diffusion coefficient (ADC) maps (Table 4-1). The mpMRIs were scored as part of standard clinical practice by experienced uro-radiologists using the five-point ordinal Likert scale of the likelihood of clinically significant prostate cancer \[69\]. The reporting radiologist was aware of the PSA level at the time of reporting. All scans had an overall Likert score and a lesion-based Likert score.
### Table 4-1 INNOVATE Scan Parameters

**Scanner: Achieva, Philips, 3T**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Coil</th>
<th>TR (ms)</th>
<th>TE (ms)</th>
<th>FA (°)</th>
<th>FOV (mm)</th>
<th>Slice Thickness (mm)</th>
<th>ACQ matrix</th>
<th>Total scan duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 TSE coronal</td>
<td>Dual</td>
<td>6128</td>
<td>100</td>
<td>90</td>
<td>180</td>
<td>3</td>
<td>300 x 290</td>
<td>05:55.4</td>
</tr>
<tr>
<td>T2 TSE axial</td>
<td>Dual</td>
<td>5407</td>
<td>100</td>
<td>90</td>
<td>180</td>
<td>3</td>
<td>300 x 290</td>
<td>05:13.6</td>
</tr>
<tr>
<td>DWI 0 150 500 1000</td>
<td>Dual</td>
<td>2753</td>
<td>80</td>
<td>90</td>
<td>220</td>
<td>5</td>
<td>168 x 169</td>
<td>05:16.5</td>
</tr>
<tr>
<td>DWI b2000</td>
<td>Dual</td>
<td>2000</td>
<td>78</td>
<td>90</td>
<td>220</td>
<td>5</td>
<td>168 x 169</td>
<td>03:40.0</td>
</tr>
<tr>
<td>DCE 2 dyn mod SENSE</td>
<td>Dual</td>
<td>5.8</td>
<td>2.8</td>
<td>90</td>
<td>180</td>
<td>3</td>
<td>140 x 177</td>
<td>00:28.9</td>
</tr>
<tr>
<td>DCE 20 dyn mod SENSE</td>
<td>Dual</td>
<td>5.8</td>
<td>2.8</td>
<td>90</td>
<td>180</td>
<td>3</td>
<td>140 x 162</td>
<td>04:14.1</td>
</tr>
</tbody>
</table>

**Scanner Siemens Avanto 1.5 T**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Coil</th>
<th>TR (ms)</th>
<th>TE (ms)</th>
<th>FA (°)</th>
<th>FOV (mm)</th>
<th>Slice Thickness (mm)</th>
<th>Matrix base</th>
<th>Matrix phase %</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 TSE Coronal</td>
<td>Body</td>
<td>5240</td>
<td>104</td>
<td>150</td>
<td>180</td>
<td>3</td>
<td>256</td>
<td>95</td>
</tr>
<tr>
<td>ep2d diffusion b1400</td>
<td>Body</td>
<td>5170</td>
<td>98</td>
<td>180</td>
<td>180</td>
<td>3</td>
<td>256</td>
<td>100</td>
</tr>
<tr>
<td>ep2d diffusion new 16</td>
<td>Body</td>
<td>2200</td>
<td>98</td>
<td>0</td>
<td>320</td>
<td>3</td>
<td>172</td>
<td>100</td>
</tr>
<tr>
<td>T1 flash 3d match VIBE</td>
<td>Body</td>
<td>2100</td>
<td>96</td>
<td>0</td>
<td>260</td>
<td>3</td>
<td>172</td>
<td>100</td>
</tr>
<tr>
<td>T1 flash 3D match VIBE</td>
<td>Body</td>
<td>10.4</td>
<td>92</td>
<td>15</td>
<td>260</td>
<td>3</td>
<td>256</td>
<td>100</td>
</tr>
</tbody>
</table>
4.2.4 **VERDICT MRI Protocol**

Men recruited in the trial also underwent VERDICT MRI in the same scanning session or within 2 weeks of the clinical mpMRI. The acquisition parameters for VERDICT MRI will be included in a later chapter.

4.2.5 **Biopsy**

The decision to offer biopsy to patients was made on the result of the mpMRI and PSA as per standard of care. For Likert scores of 4 or 5, a biopsy is recommended to the patient. If a patient has a Likert 3 score and the PSAD is below 0.15 ng/ml/ml, patients are advised that the likelihood of clinically significant cancer is low, and they are offered biopsy or surveillance. If a patient has a Likert 3 lesion and a high PSAD (>0.15 ng/ml/ml), biopsy or surveillance is advised. Men with Likert 2 lesions are discharged without biopsy.

If there was a focal lesion or regional signal abnormality identified on mpMRI, this was targeted with trans-perineal or transrectal biopsy. If there is no target, then a template mapping approach is used to sample all the regions of the prostate. Biopsy was carried out at two centres: UCLH and Barts Health. At UCLH, a transperineal approach was used whereas at Barts a transrectal approach was used.

4.2.6 **Serum PSA and PSA Density**

Patients were usually referred due to a raised serum PSA measured in primary care. The serum PSA was repeated if it was not recent (more than 3 months) before patients underwent mpMRI. The referral guidelines for serum PSA thresholds for men with suspected prostate cancer were: > 3.0 ng/mL for men aged 50-69-years-old, and > 5.0 ng/mL for men over 70-years-old [70].
PSAD was calculated by dividing serum PSA (ng/ml) by mpMRI-derived prostate volume (mL) using the prostate ellipsoid method (width x length x height x 0.52)[71].

4.2.7 **PIRADS 2.1 Rescoring**

To assess the impact of PIRADS 2.1 scoring, men who underwent biopsy were retrospectively rescored with PIRADS 2.1 by a board-certified radiologist [71].

4.2.8 **Statistical Analyses**

Non-parametric tests were used to compare differences between men with clinically significant prostate cancer and those without for PSA, PSAD and multiparametric score for the primary and secondary outcomes. The independent samples median test with Yate’s continuity correction was used to compare differences between the groups for the primary outcome. The Kruskal-Wallis test with Bonferroni correction for multiple tests was used to test differences in Gleason score groups for PSAD. Significance values were adjusted by Bonferroni correction for multiple tests.

A p-value of <0.05 was considered statistically significant for all statistical tests. Confidence intervals for median values of PSA and PSAD were derived by using bootstrapping based on 10000 samples.

4.3 **Results**

4.3.1 **Patient Demographics**

A total of 303 men were recruited in the study, who underwent mpMRI. A total of 36 men were excluded from the analysis because they refused VERDICT MRI
(n=3), had a previous negative biopsy (n=3) or had unusable VERDICT acquisitions due to artefact (n=15) or coil malfunction (n=16) (Figure 4-1).

The median age of the patients was 64 years old (range 40 to 81 years). Out of these patients 165/303, 54% underwent a biopsy. Prostate biopsy was performed a median of 28 days after mpMRI.

4.3.2 Multiparametric MRI

The maximum Likert scores for the cohort were as follows: 14% (43/303) of men had a score of 2/5, 55% had a score of 3/5 (166/303), 18% (54/303) scored 4/5 and 13% (40/303) of men scored 5/5. The decision to biopsy was associated with a higher Likert score. All patients with a Likert score of 5 were biopsied (40/40) and 91% (49/54) of patients with a score of 4 were biopsied. Biopsy was not carried out in 5 men with Likert 4 lesions because of patient preference. Nearly half of patients (47%, 76/161) with a score of 3 were biopsied (Figure 4-2).
**Figure 4-2** Proportion of patients biopsied for each Likert score. Number of patients for each component are indicated.

**Figure 4-3** Proportion of patients with clinically significant cancer subdivided by Likert score.
Figure 4-4 MRI images of 4 participants.

Each panel displays a T2W image top left, high b value image bottom left, post contrast image top right, ADC map bottom right. Red arrows indicate abnormalities.

Panel A shows MRI images of a participant scored as Likert 2 who was not biopsied.

Panel B – Right midgland was scored as 3/5 and PSAD of 0.04, not biopsied.

Panel C - participant was scored as Likert 4 in the right midgland, negative on biopsy.

Panel D – large lesion in the peripheral zone score Likert 5, positive for 4+3 disease on biopsy.
4.3.3 Biopsy

Biopsies were carried out at two specialist centres. Most men (82%, 136/165) underwent targeted trans-perineal biopsy performed at one centre, whereas 18% (29/165) of the cohort underwent targeted trans-rectal biopsy at a different specialist centre. All biopsies at UCLH and most (93%, 27/29) at Barts were performed using cognitive fusion. Two biopsies were performed using technology-assisted fusion at Barts. Biopsy was performed at a median of 28 days from imaging with most lesions biopsied within 3 months (87%, 144/165). For some patients (n=21), a biopsy was performed more than 3 months after research imaging after a period of PSA and MRI surveillance.

A total of 73/165 (44%) lesions were positive for clinically significant cancer, 16 lesions had insignificant cancer and 76 lesions were negative. Most lesions with csPCA (63%, 46/73) had 3+4 disease, followed by 4+3 (17), 4+5 (3), 3+5 (2), 4+4 (2), 5+4 (1) and 5+5 (1).

The proportion of lesions with csPCA for each Likert score was: 10/76 Likert 3, 25/49 Likert 4 and 38/40 Likert 5 men had csPCA (Figure 4-3).

4.3.4 Serum PSA

The median PSA was 5.97 (range 0.68 to 141). The distribution of PSA subdivided by Likert score is shown in Figure 4-4.
Figure 4–5 Boxplot of PSA subdivided by Likert score and decision to biopsy

Outliers are indicated by (~)

Higher mpMRI scores were associated with higher PSA. Patients were subdivided by their maximum Likert score and PSA was compared between patients who did or who did not undergo prostate biopsy. There was no statistically significant difference between these patients for Likert 3 (p=0.143) and Likert 4 (p=0.605) lesions. All Likert 5 lesions were biopsied and no Likert 2 lesions were biopsied. The distribution of PSA for patients who underwent biopsy and had clinically significant cancer or not is shown in Figure 4–5.
In addition, for patients who scored either Likert 3, 4 or 5 and underwent biopsy, there was no statistically significant difference between serum PSA of men with clinically significant cancer on biopsy compared to those without (Likert-3, p=0.734) (Likert 4, p=0.473) (Likert 5, p=0.468).

4.3.5 **PSA DENSITY**

The median PSAD was 0.137 (range 0.024-2.364, IQR 0.148). The distribution of PSAD and decision to biopsy is shown in Figure 4-6. Most men with a Likert score of 4 or 5 were biopsied regardless of PSAD. A total of 53 men who had a PSAD below 0.15 and an indeterminate mpMRI (Likert 3) score were biopsied. A total of 16 men who had PSAD below 0.08 and an indeterminate mpMRI score were biopsied.
The median PSAD for lesions with csPCa was 0.208 (range 0.05-2.388, IQR 0.232, 95% CI 0.267-0.488 95% CI) and was 0.118 (range 0.024-0.746, IQR 0.073, 95% CI 0.122-0.169) for lesions without significant cancer. This difference was statistically significant (p<0.001, independent samples median test). The distribution of PSAD subdivided by Likert score is shown in Figure 4-7.

For indeterminate mpMRI lesions, median PSAD was not significantly different (p=0.074, independent samples median test) in lesions that had csPCa (95% CI 0.144-0.285) compared to those without (95% CI 0.114-0.161). There were no significant differences in median PSAD for Likert 4 lesions either (p=0.319).

Figure 4-8 shows the correlation between PSAD and Gleason grade. PSAD increased with higher Gleason grade and there were significant differences between lesions with no cancer and lesions with 3+4 (p=0.002, K-W test) and ≥ 4+3 (p<0.001, K-W test). There was a statistically significant difference between lesions with 3+3 compared to lesions with ≥ 4+3 (p=0.016, K-W test) but not between lesions with 3+3 and 3+4 (p=0.934, K-W test). There was no significant difference between lesions with 3+4 and ≥ 4+3 (p=0.160, K-W test).
Outliers are indicated by (-), Two dashed lines represent PSA density thresholds at 0.15 and 0.08.

Outliers are indicated by (-), Two dashed lines represent PSA density thresholds at 0.15 and 0.08.

Two reference lines (dashed line) represent two PSA density thresholds. Outliers are indicated by (-)
To investigate the impact of PIRADS scoring, all men who underwent biopsy were retrospectively re-scored using PIRADS 2.1. This led to a re-classification of 52 Likert-3 lesions. A total of 42 lesions were rescored to PIRADS 2 and 10 upgraded to PIRADS 4. Two Likert 5 lesions were re-classified to PIRADS 4 (Figure 4-9). Two out of the 42 downgraded PIRADS 2 lesions were positive for clinically significant cancer.

The distribution of PSA and PSAD subdivided by PIRADS score are shown in Figures 4-10 and 4-11. There was no significant difference in PSAD in men with csPCa compared to men without indeterminate PIRADS 3 (p=0.243) but there was a difference for PIRADS 4 lesions (p=0.03). Out of 42 men rescored to PIRADS 2, 2 had csPCa (5%). Out of 24 men who scored PIRADS 3, 4 had csPCa (17%) and out of 61 men who scored PIRADS 4, 30 had csPCa (49%). Most men (37) who scored PIRADS 5 had csPCa (97%).
Figure 4-10 Proportion of patients with clinically significant cancer by PIRADS score

Figure 4-11 Distribution of PSA and PIRADS score. Outliers are indicated by (-).
Figure 4-12 Distribution of PSAD and presence of clinically significant cancer. Two PSA thresholds are indicated by dashed lines. Outliers are indicated by (-).

4.4 DISCUSSION

In summary, the clinical outcomes of the INNOVATE study show the current status of the mpMRI-directed clinical pathway. Overall, by using the risk stratification of pre-biopsy mpMRI, 46% of men avoided biopsy.

In this cohort, the most frequent Likert score ascribed was indeterminate 3/5 at 53%. Only 14% were negative (2/5) and 13% strongly positive (5/5). This distribution of Likert scores likely reflects the high prevalence of benign conditions such as atrophy or inflammation which cause abnormal signals on the components of multiparametric MRI, and therefore not many prostates have a normal signal (Likert 1). These changes can have a ‘masking’ effect which makes prostate cancer less conspicuous, reflected by more 4/5 (60/303) scores than 5/5 (40/303) [63][72].
It is clear from the cohort that the mpMRI Likert score has the strongest influence on the decision to biopsy. When the score is 4 or 5, most men (83%, 100%) undergo biopsy regardless of PSA level. However, when the score is 3/5, usually PSA density thresholds are used. In this cohort, it is also clear that these thresholds are not used rigidly, and the patient is offered a biopsy even when the PSAD is below 0.15. In fact, in men with a score of 3/5, 53 men who had a PSA density below 0.15 and 16 men who had a PSAD below 0.08 were biopsied due to patient preference. Overall 47% of men underwent biopsy in this cohort.

When the mpMRI score is strongly positive (5/5), it is highly predictive of prostate cancer on biopsy (95% have csPCa). As the mpMRI score is less positive, the percentage of csPCa found on biopsy falls. The vast majority (87%) of Likert 3 biopsies were negative for cancer or had clinically insignificant cancer. Almost half of the men who scored Likert 4 had a negative or insignificant biopsy result.

PSA was not a useful discriminator between men who had clinically significant cancer and those who did not. PSAD was a better differentiator overall. However, in men who had an indeterminate score on mpMRI, the difference in PSAD was not statistically significant. A PSAD threshold of 0.15 ng/ml/ml has been reported in the literature as a useful threshold in risk stratifying men. In this cohort, if the PSAD threshold of 0.15 ng/ml/ml is used on its own, it would result in potentially 30 men with clinically significant cancer being overlooked.

However, a combination of PSAD and Likert scores could help avoid unnecessary biopsies. If this threshold is applied to patients with a score of Likert 3, potentially 69% (52/75) men would have avoided a biopsy. However, in 4 patients clinically significant cancer would be missed. If a lower PSAD threshold is used (0.08), then only one man with clinically significant cancer might not have been diagnosed but fewer men 23% (17/75) would avoid a negative biopsy. However, at this lower threshold, 58 men would have an unnecessary negative biopsy.
The results mentioned above demonstrate the limitations of the current diagnostic pathway. More than half of mpMRI is reported as indeterminate by experienced radiologists. PSA and PSAD are routinely used to stratify patients in addition to their mpMRI score. However, this strategy can result in potentially missed clinically significant cancer. This risk is apparent in this cohort as PSA or PSAD are not used rigidly and many patients prefer to have a biopsy when the MRI is reported as indeterminate regardless of PSAD.

The relatively high number of indeterminate studies is similar to the rates reported in large multi-centre trials. There could be several reasons for this. Most readers in the context of suspected cancer, will err on the side of caution and be more sensitive than specific. This approach ensures that cancers are not missed at the expense of leading to an unnecessary biopsy. Limitations of conventional sequences due to artefact or resolution could result in a high number of indeterminate studies. Benign pathologies such as inflammation and atrophy can cause diffuse changes in the prostate which could mask small tumours. In this study, a group of 10 radiologists read the clinical MRI and therefore differences in experience may also have an impact.

The limitations of this study are that the INNOVATE study is a single tertiary referral centre and therefore has limited external validity, and generalisability. Although most men underwent trans-perineal MRI targeted, a small minority underwent systematic and combination approaches. This heterogeneity has an impact on the radiological and pathological correlation.

CONCLUSIONS

The observational cohort of the INNOVATE study represents a real-world illustration of the mpMRI-directed diagnostic pathway. This analysis sets the baseline against which the impact of VERDICT MRI can now be assessed. From this analysis, I can conclude that there is potential for improvement of the mpMRI pathway when qualitative assessment ascribes scores that are not strongly positive or indeterminate.
5 Quantitative analysis of VERDICT MRI

Author declaration

All the work in this chapter was conceived and written by me, under the supervision of Professor Shonit Punwani, Professor Daniel Alexander and Dr Eleftheria Panagiotaki. VERDICT maps were produced via an automatic fitting process using a cloud-based platform called XNAT (maintained at UCL by Baris Kanber), which integrates the VERDICT code of Dr Eleftheria Panagiotaki and her team. Dr Harriet Rodgers helped fit ROIs produced by myself to VERDICT maps to derive quantitative values. The clinical data were collected by myself and cross-checked with Dr Hayley Pye who independently also collected clinical data. Multiparametric MRI was read by a group of 10 uro-radiologists. Biopsies were performed by urologists at UCLH and Barts. Histopathology was reported by consultants at UCLH and Barts Health.

5.1 INTRODUCTION

In the previous chapter, the current risk stratification strategy for men with a suspicion of prostate cancer was analysed in a prospectively recruited cohort. This highlighted the impact of mpMRI in reducing the number of men going for biopsy in contrast to the previous pathway of using transrectal ultrasound-guided biopsy in all men with a raised PSA.

However, there remains room for improvement in risk stratification. More than half of scans were reported as indeterminate which leads to a diagnostic dilemma for clinicians. This high percentage of indeterminate studies has been seen in large studies such as PROMIS, PRECISION and meta-analyses [22,23,35]. In this uncertainty, patients are offered a biopsy based on their raised PSA, PSAD and patient preference. Most of these biopsies are negative or discover clinically insignificant disease.
Quantitative evaluation of prostate MRI has potential advantages over qualitative image analysis. For example, it may reduce the inherent subjectivity of qualitative assessment and therefore reduce inter-reader variability and lead to less indeterminate studies. It may also facilitate more accurate follow up of lesions under active surveillance or be used to monitor response to treatment.

Several biomarkers are being considered risk stratify patients\[34,73\], two of these can be derived from multiparametric MRI.

The first, lesion-ADC (Apparent diffusion coefficient), has been investigated in many research studies and demonstrates a difference between benign lesions and lesions with csPCa \[73–76\]. Despite this, prostate mpMRI interpretation remains fully qualitative [PIRADS 2.1 guidelines\[71\]]. Limitations of quantitative ADC measurements are they have only a moderate correlation with Gleason grade in the peripheral zone and a weak correlation in transition zone \[74\]; and, heterogeneity in ADC values across scanners precluding derivations of thresholds can be reliably applied for patient risk-stratification \[76\].

The second, PSA density (PSAD), as previously discussed is used routinely to select patients for biopsy particularly when the mpMRI score is indeterminate. Several PSAD thresholds have been proposed ranging from 0.08 to 0.15 (ng/ml/ml), but no consensus exists on the optimal threshold for advising men on a biopsy \[34,67\]. PSAD remains an important predictor of significant disease in non-MRI nomograms \[72\].

VERDICT MRI, as described previously, aims to improve the biological specificity of conventional diffusion-weighted imaging through its optimised design for prostate microarchitecture. VERDICT aims to more precisely depict histological changes associated with cancer and thereby improve patient risk stratification.
This chapter aims to compare quantitative VERDICT parameters: fractional intracellular volume (FIC), vascular fraction (FVASC), extracellular extravascular space (FEES) to apparent diffusion coefficient (ADC) and PSA density (PSAD) in differentiating lesions with clinically significant cancer (csPCa) from lesions with benign or insignificant cancer on targeted biopsy. Further analysis will investigate the correlation of these parameters to Gleason Grade.

### 5.1.1 AIMS

i) Test FIC, FVASC and FEES in differentiating clinically significant cancer

ii) To compare VERDICT parameters to PSAD and ADC from mpMRI in differentiating clinically significant cancer

iii) To test whether VERDICT parameters correlate with Gleason Grade

### 5.1.2 NULL HYPOTHESES

i) The distribution of FIC/FEES/FVASC is the same across clinically significant cancer and benign/non-clinically significant cancer

ii) The distribution of PSA density/ADC is the same for patients with clinically significant cancer and benign/non-clinically significant cancer

iii) The distribution between VERDICT or mpMRI parameters is the same for Gleason Grades

### 5.2 METHODS

#### 5.2.1 PARTICIPANTS
Men who underwent targeted biopsy of abnormalities identified on the clinical mpMRI were selected for this analysis from the INNOVATE study described previously. This included men recruited from both centres: University College London Hospital (UCLH) and Barts Health, were included.

Exclusion criteria were the same as those applied in the previous chapter. Men who had a history of previous prostate biopsy were excluded. Men who had unusable diffusion data due to artefact or equipment failure were also excluded.

5.2.2 STUDY DESIGN

The following clinical data were collected: age, PSA at time of referral, mpMRI Likert score, prostate volume, pictorial biopsy target maps and biopsy reports. The index lesion was defined as the highest-scoring lesion identified on mpMRI with Likert scores (3-5). Clinically significant cancer was defined as the presence of a single biopsy core containing Gleason $\geq 3 + 4/ $ Gleason grade group 2 or higher as defined by the International Society of Urological Pathology [77]. Lesion Gleason score was defined as the predominant pattern in the biopsy cores taken at the corresponding target.

Two groups were defined for the primary outcome: 1) no cancer/insignificant cancer (non-significant histology) and 2) clinically significant cancer (csPCa). A subgroup analysis of indeterminate (Likert 3) lesions was also performed. For Gleason score correlation, four groups were defined: no cancer, 3+3, 3+4, $\geq 4 + 3$. The $\geq 4 + 3$ group combined the higher Gleason grades to have comparable numbers to other groups.
5.2.3 **Multiparametric MRI and VERDICT MRI**

Multiparametric MRI was performed at 3T or 1.5T compliant with Prostate Imaging Reporting and Data System version 2.1 (PI-RADS v2.1) standards at University College London Hospital[78]. The protocol for multiparametric MRI has been mentioned earlier. VERDICT MRI was performed at 3T and used b values of 0-3000 s/mm²).

The VERDICT MRI protocol requires a 32-channel cardiac coil placed over the pelvis. The sequence uses b values of 90-3000 s/mm² in three orthogonal directions. The acquired data is normalized with a b = 0 image for every echo time (TE) to avoid T₂ dependence. The scanning parameters for VERDICT MRI are provided in Table 5-1.
### Table 5-1 VERDICT scan parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Phillips Achieva (3T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR scanner</td>
<td>32 channel Cardiac coil</td>
</tr>
<tr>
<td>Sequence</td>
<td>DWI SE EPI single shot</td>
</tr>
<tr>
<td>Field of View (mm)</td>
<td>220</td>
</tr>
<tr>
<td>Number of slices</td>
<td>14</td>
</tr>
<tr>
<td>Slice thickness (mm)</td>
<td>5</td>
</tr>
<tr>
<td>Phase encoding direction</td>
<td>AP</td>
</tr>
<tr>
<td>Reconstructed matrix</td>
<td>176 x 176</td>
</tr>
<tr>
<td>Reconstructed pixel size (mm)</td>
<td>1.25</td>
</tr>
<tr>
<td>Water fat shift (WFS[pix])/Bandwidth(Hz)</td>
<td>49.09/8.8</td>
</tr>
<tr>
<td><strong>b-values</strong></td>
<td></td>
</tr>
<tr>
<td>Repetition time (TR) range, actual (ms)</td>
<td>3349-10000, 2260</td>
</tr>
<tr>
<td>Echo time (TE) (ms)</td>
<td>80</td>
</tr>
<tr>
<td>DELTA/delta (ms)</td>
<td>38.8/18.9</td>
</tr>
<tr>
<td>Number of signal averages</td>
<td>6</td>
</tr>
<tr>
<td><strong>b-values</strong></td>
<td></td>
</tr>
<tr>
<td>TR range, actual (ms)</td>
<td>2000-10000, 3897</td>
</tr>
<tr>
<td>TE(ms)</td>
<td>67</td>
</tr>
<tr>
<td>DELTA/delta (ms)</td>
<td>92.5/12.8</td>
</tr>
<tr>
<td>Number of signal averages</td>
<td>6</td>
</tr>
<tr>
<td><strong>b-values</strong></td>
<td></td>
</tr>
<tr>
<td>TR range, actual (ms)</td>
<td>2482-10000, 2482</td>
</tr>
<tr>
<td>TE(ms)</td>
<td>65</td>
</tr>
<tr>
<td>DELTA/delta (ms)</td>
<td>31.5, 11.4</td>
</tr>
<tr>
<td>Number of signal averages</td>
<td>6</td>
</tr>
<tr>
<td><strong>b-values</strong></td>
<td></td>
</tr>
<tr>
<td>TR range, actual (ms)</td>
<td>2482-10000, 2482</td>
</tr>
<tr>
<td>TE(ms)</td>
<td>65</td>
</tr>
<tr>
<td>DELTA/delta (ms)</td>
<td>25.8/3.9</td>
</tr>
<tr>
<td>Number of signal averages</td>
<td>4</td>
</tr>
<tr>
<td>Acquisition Time (minutes: seconds)</td>
<td>10.95</td>
</tr>
</tbody>
</table>
5.2.4 BIOPSY

The decision to offer biopsy was based on the standard of care at University College London Hospital and Barts Health, which is based on mpMRI Likert score and PSA value. For Likert scores of 4 or 5, a biopsy is highly recommended to the patient. If a patient has a Likert 3 score and the PSAD is below 0.15 ng/ml/ml, patients are advised that the likelihood of clinically significant cancer is low, but both biopsy and surveillance is offered. If a patient has a Likert 3 lesion and a high PSAD (>0.15 ng/ml/ml), a biopsy is recommended however the patient can also opt for surveillance. Surveillance consists of 6-monthly PSA levels and a mpMRI in a year.

For men who proceeded to biopsy, focal lesions, or regions of abnormal signal scored Likert 3 or above on mpMRI were targeted with trans-perineal or transrectal biopsy at both centres.

5.2.5 ANALYSIS

The index MRI lesion was selected as the highest scoring lesion based on two reads of the mpMRI by experienced uro-radiologists (>10 years’ experience). A Region of Interest (ROI) was drawn manually on FIC, FVASC, FEES and ADC maps. The ROIs were then used by another researcher to derive mean parameter values on MATLAB (The Mathworks Inc., Natick, Massachusetts) (Figure 5-1). PSA density was calculated by dividing the referral PSA by MRI-calculated prostate volume using the ellipsoid method.

Lesions were also retrospectively re-scored using the Prostate Imaging Reporting and Data system 2.1 (PIRADS 2.1) by a board-certified radiologist blinded to histology, to compare the distribution of VERDICT parameters (FIC, FVASC and FEES), lesion-ADC, and PSAD [79].
Shapiro-Wilk test was used to test the data for normal distribution. Non-parametric tests were used to compare differences between groups for the primary and secondary outcomes. For the primary outcome, differences in median values of PSAD, lesion-ADC and lesion-FIC between the two groups (csPCa vs non-significant histology) was analysed by the independent samples median test with Yate’s continuity correction. The Kruskal-Wallis test with Bonferroni correction for multiple tests was used to test differences in the four Gleason score groups. Significance values were adjusted by Bonferroni correction for multiple tests. Confidence intervals for median values of the parameters were derived by using bootstrapping based on 10000 samples.

Sub-group analyses were carried out for men with Likert 3 lesions, Likert 4 lesions, PIRADS 2.1 re-scoring and men who had mpMRI at 1.5T or 3T. Likert 5 lesions were not included in the subgroup analysis because in this cohort almost all Likert 5 lesions (95%) were positive for csPCa. ADC values can vary depending on field strength and therefore a subgroup analysis was carried out to determine this effect[80,81].

Statistical analyses were performed using SPSS v.26 (IBM Corp., Armonk, NY, USA) and GraphPad Prism v.9 (GraphPad Software, La Jolla, California, USA). The p-value threshold for significance was set at p < 0.05.
Figure 5-1 MR Images from three men in the INNOVATE study.

The left-most image is an axial T2-weighted image (T2W), the centre image is a VERDICT intracellular volume fraction map (FIC), and the right-most image is an Apparent Diffusion Coefficient (ADC) map. A Region of interest (ROI) delineates the biopsied lesions on FIC and ADC maps (red ellipse) and arrow on the T2W image.

Panel (a) 58-year-old man with an MRI-lesion in the anterior basal prostate positive for prostate cancer..

Panel (b) 68-year-old man with an MRI lesion in the right apical posterolateral peripheral zone which was negative for prostate cancer on biopsy.

Panel (c) 64-year-old man with an MRI lesion in the right basal posterolateral peripheral zone which was negative for prostate cancer on biopsy.
5.3 RESULTS

5.3.1 PATIENTS, PSA AND PROSTATE VOLUME

A total of 303 biopsy naïve men underwent mpMRI and VERDICT MRI. The median age of participants was 64 years old (range 47 to 81 years). Out of these patients, 165 underwent targeted biopsy. The median PSA was 6.48 ng/ml (range 0.83 to 141). The median prostate volume was 46 ml (range 12–150).

5.3.2 MULTIPARAMETRIC MRI LESIONS

The Likert scores for the biopsied lesions were as follows: 46% (76/165) scored Likert 3/5, 30% (49/165) scored Likert 4/5 and 24% (40/165) scored 5/5. Most index lesions were in the peripheral zone (PZ) (78%, 129/165), followed by the transition zone (15%, 24/165) and rarely in the central zone (1%, 2/165). Some lesions spanned both the PZ and TZ (5%, 9/165). Most men had more than one mpMRI lesion (scored Likert 3 or higher) (70%, 115/165).

Retrospective PIRADS scoring of lesions led to a re-classification of 52 Likert-3 lesions. A total of 42 lesions were rescored to PIRADS 2 and 10 upgraded to PIRADS 4. Two Likert 5 lesions were re-classified to PIRADS 4.

5.3.3 BIOPSY

The biopsy results for this cohort have been described in the previous chapter. In summary, a total of 73/165 (44%) lesions were positive for clinically significant cancer, 16 lesions had insignificant cancer and 76 lesions were negative. Most lesions with csPCa (63%, 46/73) had 3+4 disease, followed by 4+3 (17), 4+5 (3), 3+5 (2), 4+4 (2), 5+4 (1) and 5+5 (1).
The proportion of lesions with csPCa for each Likert score was: 10/76 Likert 3, 25/49 Likert 4 and 38/40 Likert 5 men had csPCa. For PIRADS re-scored lesions, the proportion was 2/42 PIRADS 2, 4/24 PIRADS 3, 30/61, 37/38 PIRADS 5.

5.3.4 **PSA DENSITY**

PSA-density results were stated in the previous chapter. The distribution of PSA density subdivided by Likert score is shown in Figure 5-2.

![Figure 5-2](image)

**Figure 5-2** Distribution of PSA density by Likert and PIRADS score and presence of clinically significant cancer.

Statistical significance is indicated by ‘*’. Outliers are denoted by (-).
For indeterminate mpMRI lesions (Likert 3), median PSAD was not significantly different (p=0.074, independent samples median test, p=0.153) in lesions that had csPCa (95% CI 0.144-0.285) compared to those without (95% CI 0.114-0.161). For Likert 4 lesions, there was no significant difference between the two groups in median PSAD (p=0.319). For PIRADS 2 and 3 lesions, there was no significant difference in median PSAD (p=0.469, p=0.590). There was a significant difference in median PSAD (p=0.03) for PIRADS 4 lesions.

Figure 5-3 shows the correlation between PSAD and Gleason grade. PSAD increased with higher Gleason grade and there were significant differences between lesions with no cancer and lesions with 3+4 (p=0.002, K-W test) and ≥ 4+3 (p<0.001, K-W test). There was a statistically significant difference between lesions with 3+3 compared to lesions with ≥ 4+3 (p=0.016, K-W test) but not between lesions with 3+3 and 3+4 (p=0.934, K-W test). There was no significant difference between lesions with 3+4 and ≥ 4+3 (p=0.160, K-W test).
5.3.5 ADC

The median lesion-ADC for lesions that had significant cancer was 0.831 x 10^{-3} \text{ mm}^2/\text{s} (range 0.113-1.596, IQR 0.286, 95% CI 0.765-0.885 x 10^{-3} \text{ mm}^2/\text{s}) and 1.170 x 10^{-3} \text{ mm}^2/\text{s} (range 0.587-2.110, IQR 0.358, 95% CI 1.129-1.240 x 10^{-3} \text{ mm}^2/\text{s}) in lesions without significant cancer. This difference in medians (p<0.001) and distribution was statistically significant (p<0.001).

For Likert-3 lesions, median lesion-ADC was not significantly different (p=0.090) in lesions that had csPCa compared to those without (Figure 5-4).

There was a significant difference in median lesion-ADC (p=0.032) for lesions scored Likert-4 for both groups.

For PIRADS 2 and 3 lesions, there was no significant difference in median ADC (p=0.469, p=0.93) or data distribution (p=0.509, p=0.181). There was a significant difference in median ADC (p<0.001) and data distribution (p<0.001) for PIRADS 4 lesions.
Figure 5–4 Distribution of lesion-ADC subdivided by Likert and PIRADS Score and presence of clinically significant cancer.

Statistical significance is indicated by ‘*’. Outliers are denoted by (-).

Figure 5–5 Lesion-ADC and Gleason score correlation.

Statistical significance is indicated by ‘*’. Outliers are denoted by (-).

Lesion-ADC was lower for lesions with higher Gleason grade tumours (Figure 5–5). There was a statistically significant difference in lesion-ADC between lesions without cancer and lesions with 3+4 (p=0.001, K-W test) or ≥ 4+3 disease (p=0.032, K-W test). There was a statistically significant difference between
lesions with 3+3 compared to lesions with 3+4 (p=0.045, K-W test) but not between 3+3 and ≥ 4+3 (p=0.114, K-W test). There was no statistically significant difference between lesion-ADC in lesions with 3+4 compared to lesions with ≥ 4+3 (p=0.99).

5.3.5.1 1.5 T and 3T mpMRI subgroup

A subgroup analysis of patients who had clinical DWI acquisitions on the same scanner (3T) as VERDICT (53%, 87/165) was performed (Figure 5-6). The results showed similar results to the rest of the cohort. There was no significant difference in median lesion-ADC for lesions with csPCa compared to those without Likert 3 lesions (p=0.06). However, for Likert 4 lesions, there was a significant difference in median ADC for the two groups (p= 0.005).

The correlation with Gleason grade was also similar compared to the entire cohort (Figure 5-7) with a statistically significant difference between 3+3 and 3+4 (p=0.048, K-W test), but not between 3+3 and ≥ 4+3 (p=0.192, K-W test) or 3+4 and ≥ 4+3 (p=0.767, K-W test).

For the 1.5T subgroup, there were no differences in median ADC for lesions with csPCa compared to those without for Likert 3 (p=0.063) or 4 (p=0.10) lesions. There was a difference between negative lesions and lesions with 3+4 (p<0.001) and lesions with ≥ 4+3 disease (p<0.001) but not between negative lesions and 3+3 (p>0.99). There was also a difference between 3+3 and ≥ 4+3 disease (p=0.001) and 3+3 and 3+4 (p=0.013). There was no difference between 3+4 and ≥ 4+3 disease (p>0.99).
Figure 5-6 Lesion ADC distribution subdivide by Likert score and presence of Clinically Significant cancer for men who underwent mpMRI at 3T. Statistical significance is indicated by ‘*’. Outliers are denoted by (−).

Figure 5-7 Lesion ADC distribution and Gleason grade for men who underwent mpMRI at 3T. Statistical significance is indicated by ‘*’. Outliers are denoted by (−).
Figure 5-8 Lesion ADC distribution subdivide by Likert score and presence of Clinically Significant cancer for men who underwent mpMRI at 1.5T.

Outliers are denoted by (-).

Figure 5-9 Lesion ADC distribution and Gleason grade for men who underwent mpMRI at 3T.

Statistical significance is indicated by ‘*’. Outliers are denoted by (-).
5.3.6 FIC

Median lesion-FIC for lesions with csPCa was 0.610 (range 0.010-0.885, IQR 0.166, 95% CI 0.563-0.631) and in lesions without significant cancer was 0.217 (range 0.052-0.506, IQR 0.190, 95% CI 0.217-0.266). This difference was statistically significant (p<0.001).

![Distribution of FIC by Likert and PIRADS score and presence of clinically significant cancer.](image)

Outliers are denoted by (-). * signifies statistical significance.

For indeterminate Likert lesions, there was a statistically significant difference (p<0.001) in median lesion-FIC for lesions with csPCa (95% CI 0.465-0.590) compared to lesions without (95% CI 0.194-0.248) (Figure 5-8). There was also a significant difference in median lesion-FIC (p<0.001) and for lesions scored 4/5 (Figure 5-8).

For PIRADS 2, 3 and 4 lesions, there was a significant difference in the two groups (p=0.044, p=0.005, p<0.001).
Lesion-FIC increased with increasing Gleason grade (Figure 5-9). There was a statistically significant difference in lesion-FIC between lesions with no cancer and lesions with 3+4 disease (p<0.001, K-W test) and ≥ 4+3 PCa (p<0.001, K-W test) (Figure 5-9). There was a statistically significant difference between lesions with 3+3 and lesions with 3+4 (p<0.001, K-W test) and ≥ 4+3 PCa (p<0.001, K-W test).

![FIC and Lesion Gleason Grade](image)

**Figure 5-11** Lesion FIC and Gleason Grade correlation.
Statistical significance is indicated by ‘*’. Outliers are denoted by (−).

### 5.3.7 FEES

Median lesion-FEES for lesions with csPCa was 0.305 (range 0.118-0.871, IQR 0.133, 95% CI 0.288-0.334) and in lesions without significant cancer was 0.554 (range 0.077-1.119, IQR 0.176, 95% CI 0.509-0.564). This difference was statistically significant (p<0.001).
For indeterminate lesions, there was a statistically significant difference ($p=0.018$) in median lesion-FEES for lesions with csPCa (95% CI 0.305-0.433) compared to lesions without (95% CI 0.525-0.588) (Figure 5-10). There was also a significant difference in median lesion-FIC ($p<0.001$) for lesions scored 4/5.

For PIRADS 2 and 3 lesions, there was no significant difference in median FEES ($p=0.053$, $P=0.093$). There was a significant difference in median FEES ($p=0.007$) for PIRADS 4 lesions.

Lesion-FEES decreased with increasing Gleason grade (Figure 5-11). There was a statistically significant difference in lesion-FIC between lesions with no cancer and lesions with 3+4 disease ($p<0.001$, K-W test) and $\geq 4+3$ PCA ($p<0.001$, K-W test) but not 3+3 ($p=0.690$, K-W test). There was a statistically significant difference between lesions with 3+3 and lesions with 3+4 ($p<0.001$, K-W test) and $\geq 4+3$ PCA ($p<0.001$, K-W test). There was no statistically significant difference between lesions with 3+4 and $\geq 4+3$ PCA ($p=0.99$, K-W test).
Figure 5-13 Lesion FEES and Gleason grade correlation.
Statistical significance is indicated by ‘*’. Outliers are denoted by (-).

5.3.8  *FVASC*
Median lesion-FVASC for lesions with csPCa was 0.256 (range 0.05-0.844, IQR 0.152, 95% CI 0.187-0.238) and in lesions without significant cancer was 0.3 (range 0.072-0.824, IQR 0.122, 95% CI 0.301-0.5). This difference was statistically significant (p<0.001).
Figure 5-14 Distribution of lesion-FVASC subdivided by Likert and PIRADS Score and presence of clinically significant cancer.

Statistical significance is indicated by ‘*’. Outliers are denoted by (–).

For indeterminate lesions, there was no statistically significant difference (p=0.098) in median lesion-FVASC for lesions with csPCa (95% CI 0.219-0.286) compared to lesions without (95% CI 0.294-0.361) (Figure 5-12). There was no significant difference in median lesion-FVASC (p=0.108) for lesions scored 4/5.

For PIRADS 2 or 3 lesions, there was no significant difference in median FVASC (p=0.469, P=0.590) or data distribution (p=0.108, P=0.266) There was a significant difference in median FVASC (p=0.041) for PIRADS 4 lesions.

Lesion-FVASC appeared to decrease with increasing Gleason grade (Figure 5-13). There was a statistically significant difference in lesion-FVASC between lesions with no cancer and lesions with 3+4 disease (p<0.001, K-W test) and ≥ 4+3 PCa (p<0.001, K-W test) but not 3+3 (p=0.055). There was no statistically significant difference between lesions with 3+3 and lesions with 3+4 (p=0.514, K-W test) and ≥ 4+3 PCa (p=0.329, K-W test).
Figure 5.12 FVASC and Gleason grade.

ROC analysis was performed to compare the classification ability of PSAD, ADC, FIC, FVASC and FEES and lesion-FIC in differentiating lesions with clinically significant cancer (Figure 5-14). The area under the curve (AUC) was highest for lesion-FIC at 0.960 (0.925-0.995, 95% CI), followed by FEES at 0.896 (0.847-0.945, 95% CI), ADC at 0.846 (0.785-0.907, 95% CI), FVASC at 0.787 (0.716-0.858, 95% CI) and PSAD at 0.739 (0.663-0.816, 95% CI).

The AUC was significantly higher for lesion-FIC at 0.960 (95% CI: 0.925-0.995) compared to ADC at 0.846 (p=0.016, 95% CI: 0.785-0.907).

The area under the curve for the 3T subgroup for ADC was 0.834 (95% CI: 0.745-0.924) and for lesion FIC was 0.954 (95% CI: 0.905-1.0). For the 1.5T subgroup, the AUC for ADC was 0.894 (95% CI: 0.826-0.961) and for FIC was 0.965 (95% CI: 0.912-1.0).
The optimal sensitivity and specificity indicated by the highest value of Youden’s J statistic was calculated using the formula ($J = \text{sensitivity} + \text{specificity} - 1$). This value for each biomarker was as follows: PSAD of 0.177 (sensitivity 59%, specificity 83%), lesion-ADC of $0.998 \times 10^{-3}$ (sensitivity 83%, specificity 77%), lesion-FIC of 0.411 (sensitivity 95%, specificity 90%), FEES of 0.459 (sensitivity 90%, specificity 76%) and FVASC of 0.274 (sensitivity 64%, specificity 36%).

Thresholds based on the highest Youden’s statistic were retrospectively applied in men with Likert-3 or 4 lesions ($n=125$); of which 35/125 (28%) lesions were positive for csPCa and 90/125 (72%) lesions were negative for csPCa.

If a lesion-FIC threshold of 0.411 is applied to this sub-group, so that any man with a lesion-FIC of 0.411 or below is not biopsied, 82/90 (91%) men could have avoided a non-significant biopsy and 4/35 (11%) men would have been missed with csPCa. If a FEES threshold of 0.459 is applied to this sub-group, 68/90 (76%) men could have avoided a non-significant biopsy and 5/35 men would be missed with

---

**Figure 5-16 ROC curves for FIC, PSAD, ADC, FEES, FVASC**
csPCa. If the lesion-ADC threshold of $0.998 \times 10^{-3} \text{ mm}^2/\text{s}$ was applied, 70/90 (78%) men could have avoided a non-significant biopsy and 8/35 (23%) men with csPCa would have been missed. If a PSAD threshold of 0.177 was applied, 75/125 men would have avoided a non-significant biopsy and 19 men with csPCa would have been missed. If a PIRADS threshold is applied to this subgroup so that any man with a PIRADS score of 2/5 is not biopsied, then 2/35 (6%) men with csPCa would be missed, and 40/90 (44%) would avoid a non-significant biopsy.

Similarly, for PIRADS scoring, thresholds based on the highest Youden’s statistic were retrospectively applied in men with PIRADS–3 or 4 lesions ($n=85$); of which 34/85 (40%) lesions were positive for csPCa and 51/85 (60%) lesions were negative for csPCa.

If a lesion-FIC threshold of 0.411 is applied to this sub-group, so that any man with a lesion-FIC of 0.411 or below is not biopsied, 47/51 (92%) men could have avoided a non-significant biopsy and 9/34 (26%) men would have been missed with csPCa. If a FEES threshold of 0.459 is applied to this sub-group, 40/51 (78%) men could have avoided a non-significant biopsy and 10/34 (29%) men would be missed with csPCa. If the lesion-ADC threshold of $0.998 \times 10^{-3} \text{ mm}^2/\text{s}$ was applied, 36/51 (71%) men could have avoided a non-significant biopsy and 8/34 (24%) men with csPCa would have been missed. If a PSAD threshold of 0.177 was applied, 47/51 (92%) men would have avoided a non-significant biopsy and 21/34 (62%) men with csPCa would have been missed.

5.4 DISCUSSION

In this chapter, the quantitative values of three parameters derived from VERDICT MRI were compared to parameters derived from multiparametric MRI in differentiating lesions that have clinically significant cancer on targeted biopsy in biopsy naïve men. In addition, the correlation of parameters with Gleason grade was also investigated. Two VERDICT derived parameters; FIC and FEES
provided an excellent classification of lesions with csPCa and demonstrated more marked differences between Gleason grades compared to ADC and PSAD. For indeterminate lesions ascribed by two experienced uro-radiologists, only FIC and FEES were significantly higher in lesions with csPCa compared to those without. Thresholds for these parameters were derived to better stratify men who have Likert 3 (indeterminate) or Likert 4 lesions. The best performing threshold when retrospectively applied to the cohort was for FIC, which could potentially avoid 91% of biopsies in men who had insignificant histology on biopsy at the cost of missing 11% of men with csPCa. The use of the term ‘missing’ should not be misinterpreted as it only indicates that cancer is not detected at a specifically given timepoint. In our institution and at others, it is standard practice to follow-up men with elevated PSAD and indeterminate mpMRI findings that do not undergo biopsy (either by choice or because of other clinical factors); with potential future biopsy being re-offered if a lesion declares itself on follow-up mpMRI or there is a significant PSAD rise.

FVASC was shown to not be a useful classifier of csPCa. This could either be due to the limitation of the VERDICT model in assessing vascularity or that vascularity is not the most significantly perturbed tissue fraction in csPCa compared to intracellular volume fraction and extra-cellular and extra-vascular fraction. Further work needs to correlate histological vascular fraction in tumours and VERDICT estimated vascular fraction.

The performance of PSAD and ADC in lesion stratification was comparable to reported metrics in previous studies [63,82–84].

Combining a parameter threshold with Likert scoring was better than applying the threshold to PIRADS stratified lesions because fewer men would be potentially missed with csPCa. For instance, in the subset of men with PIRADS 3 and 4 lesions, if an FIC threshold of 0.411 is applied, 9/34 (26%) of men would be missed with csPCa compared to 4/35 (11%) for Likert 3 and 4 lesions. A similar number would avoid biopsy for PIRADS (47/51, 92%) and Likert (80/82, 91%). If we
account for lesions that were reclassified from Likert 3 to PIRADS 2 (52) (which would not be biopsied), a similar percentage of men would avoid an overall unnecessary biopsy (89/93, 96%) compared to Likert + FIC threshold (82/90, 91%). However, more men with csPCa might be overlooked (11/36, 31%) with PIRADS compared to Likert (4/35, 11%). This finding is explained by the fact that Likert scoring is more sensitive than PIRADS 2.1 and captures those men with diffuse changes that have csPCa whereas in PIRADS 2.1 these are classified as PIRADS 2 lesions.

In our centre and others in the UK, scoring of mpMRI studies is performed using the Likert system as recommended by UK NICE guidelines[48]. Scoring of lesions with PIRADS 2.1 was performed retrospectively within this study and did not influence biopsy decisions. Nonetheless, lesion-FIC was able to differentiate between lesions with and without csPCa in PIRADS 3/4 scored groups. This rescored analysis implies the potential generalisability of our results to centres that exclusively use PIRADS scoring.

Overall VERDICT parameters were better classifiers compared to ADC and PSAD. This difference is likely due to the design of the VERDICT model which is based on prostate histology compared to the simpler apparent diffusion coefficient model. FIC has been shown to directly correlate with cellular density and is, therefore, more likely to correlate with Gleason grade than ADC as demonstrated in this study. PSA is known to be non-specific to cancer and it is not surprising that PSAD can also be elevated in benign conditions and therefore not as specific. Greater biological specificity of VERDICT parameters could help in deciding whether to biopsy indeterminate lesions. Furthermore, the ability to predict Gleason grade before biopsy could help with prognostication and more timely management of higher-grade disease.

This imaging for the cohort was performed at a single specialist centre and this limits its generalisability, particularly for centres in the early phase of mpMRI adoption. The clinical decision to biopsy was influenced by PSA and mpMRI score
which introduces selection bias to the lesions that were biopsied. However, this bias probably causes an underestimation of the utility of the biomarkers examined as lesions were already pre-stratified. This study design allowed comparison with the ‘real world’ diagnostic pathway and is ethically sound when assessing novel biomarkers. The scoring of lesions with PIRADS 2.1 was retrospective and biopsy decisions were not based on PIRADS scoring. This limits the generalisability of our results to centres that exclusively use PIRADS scoring, although the agreement between both systems has been shown in the literature [85,86]. ADC maps were derived from both 1.5T and 3T acquisitions whereas VERDICT parameters were only derived from 3T and therefore the cohort may not be representative. Sub-group analysis of ADC derived from 1.5T and 3T acquisitions showed similar trends compared to the entire cohort. A similar finding has been reported in the ADC of upper abdominal organs showed comparable values at 1.5T and 3T [81].

Conclusions

VERDICT derived parameters provided a better classification of csPCa and demonstrated more marked differences between Gleason grades compared to ADC and PSAD. Quantitative assessment of lesion FIC and FEES offers men with indeterminate mpMRI lesions (Likert 3) or not strongly positive mpMRI scores (Likert 4) the opportunity to safely avoid unnecessary biopsy. Prospective application of lesion-FIC thresholds before men are biopsied is the next step in validation.
6 Inter-reader agreement, image quality and qualitative assessment of VERDICT MRI in a multi-reader study

Author declaration

All the work in this chapter was conceived and written by me, under the supervision of Professor Shonit Punwani. VERDICT maps were produced via an automatic fitting process using a cloud-based platform called XNAT (maintained at UCL by Baris Kanber), which integrates the VERDICT code of DrEleftheria Panagiotaki. The two readers who assessed image quality and scored the datasets were Dr Louise Dickinson and Dr Francesco Giganti. I conducted the training session for the two readers as mentioned below. Biopsies were performed by urologists at UCLH and Barts. Histopathology was reported by consultants at UCLH and Barts Health.

6.1 INTRODUCTION

In the previous chapter, I investigated whether biomarkers from VERDICT MRI could help risk-stratify lesions that have been already identified on multiparametric MRI. Fractional intracellular volume (FIC) showed the most potential value in risk stratifying positive lesions which are scored as indeterminate (Likert 3) or likely to have prostate cancer (Likert 4). In this chapter, I investigate whether FIC maps from VERDICT could be evaluated qualitatively by radiologists and compare their image quality, inter-reader agreement to apparent diffusion maps (ADC) which are currently used.

In current clinical practice, multiparametric MRI is qualitatively evaluated by radiologists for the presence of clinically significant cancer. Focal lesions or abnormal regions of the prostate are scored on the likelihood of the presence of clinically significant cancer using either one of two scoring systems, which have been described earlier. In brief, the Prostate Imaging Reporting and Data System (PI-RADS) guidelines specify a different dominant sequence for the peripheral and transition zone. For abnormalities in the peripheral zone, DWI images are the
dominant sequence whereas, for lesions in the transition zone, the T2W sequence is the dominant sequence [71,87]. DCE images can upgrade the score for lesions in the PZ if they demonstrate early enhancement. DWI images can upgrade the score for lesions in the TZ. The alternative to this scoring system is a subjective ‘Likert’ score which ascribes a score of 1 to 5 based on a radiologist’s assessment of the mpMRI and suspicion of prostate cancer [31,86]. This scoring system evaluates the same features as PIRADS but does not strictly follow a rigid set of rules for scoring specific zones; it has been assessed for inter-observer variability and compared with PIRADS with comparable results [88,89]. PIRADS does not specifically incorporate FIC maps, but the guidelines can be applied similarly to ADC maps which are included in PIRADS. Both scoring systems were evaluated in this study as the choice of scoring system could impact inter-reader variability.

Image quality has an important impact on the diagnostic accuracy of MRI and its interpretation by radiologists in the detection of prostate cancer. Low diagnostic quality can reduce accuracy and reduce confidence in diagnosis leading to indeterminate scores [90,91]. Therefore, it is important to assess image quality before using a scoring system. Recently a scoring system has been developed called Prostate Imaging Quality (PI-QUAL) which aims to assess the quality of mpMRI against objective technical criteria and subjective criteria. The PIQUAL score is based on a 1-to-5 Likert scale where the lowest score of 1 indicates all sequences are below the standard of diagnostic quality and the highest score (5) implies that all sequences are of optimal quality [92]. The scoring system is given in Figure 6-1. For each sequence (T2-W, DWI and DCE), there are specific criteria that need to be assessed. However, there is no specified threshold at which the sequence becomes non-diagnostic. Quality can also be assessed by a simpler Likert scale used in published studies which are more subjective [93,94]. This scale is also a 5 point Likert scale which can be applied to each image type. The definitions are given in the subsequent section.

In this study, two readers qualitatively assessed a multiparametric protocol which either had an ADC map (ADC reads) or FIC map (FIC reads) for each participant and ascribed both Likert and PIRADS 2.1 scores. They assessed the quality of
images using a Likert scale (image quality) out of 5 for individual image types \[95\] and the overall dataset using a recently published system called Prostate Imaging Quality system (PI-QUAL) \[96\].

The primary objective is to determine the inter-reader agreement and image quality of FIC and ADC image types. I hypothesise that inter-reader agreement and image quality will be the same for both image types when assessed by two readers. The secondary objective is whether the number of false positives and true positives is different for FIC vs ADC reads. The hypothesis is that there will be a 10% reduction in false positives for FIC reads vs ADC reads and a similar percentage of true positives. Time taken to read FIC or ADC will also be determined.

6.1.1 AIMS

i) To evaluate image quality for FIC and ADC maps

ii) To evaluate inter-observer agreement for FIC vs ADC individual maps and whole datasets

iii) Test the number of false positives and true positives for FIC vs ADC individual maps and overall datasets

6.1.2 NULL HYPOTHESES

i) ADC maps are rated as higher quality compared to FIC maps

ii) Inter-observer agreement is higher for ADC maps compared to FIC maps

iii) The number of false positives and true positives is the same for FIC and ADC maps.
### Prostate Imaging QUALity control (PI-QUAL) scoring sheet

<table>
<thead>
<tr>
<th>PI-QUAL score</th>
<th>Criteria</th>
<th>Clinical implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All mpMRI sequences are below the minimum standard of diagnostic quality</td>
<td>It is NOT possible to rule out all significant lesions.</td>
</tr>
<tr>
<td>2</td>
<td>Only one mpMRI sequence is of acceptable diagnostic quality</td>
<td>It is NOT possible to rule out all significant lesions.</td>
</tr>
<tr>
<td>3</td>
<td>At least two mpMRI sequences taken together are of diagnostic quality</td>
<td>It is possible to rule out all significant lesions.</td>
</tr>
<tr>
<td>4</td>
<td>Two or more mpMRI sequences are independently of diagnostic quality</td>
<td>It is possible to rule out all significant lesions.</td>
</tr>
<tr>
<td>5</td>
<td>All mpMRI sequences are of optimal diagnostic quality</td>
<td>It is possible to rule out all significant lesions.</td>
</tr>
</tbody>
</table>

*Therefore reports should not include PI-RADS or UIRad scores

#### T2-WI

**Technical parameters**
- Axial plane
- Sagittal or coronal plane
- Adequate field of view
- Adequate in plane resolution
- Adequate slice thickness
- Z-axis correctly positioned

**Visual assessment**
- Capsule clearly delineated
- Seminal vesicles clearly delineated
- Ejaculatory ducts clearly delineated
- Neurovascular bundles clearly delineated
- Sphincter muscles clearly delineated
- Absence of artifacts (e.g. movement)

Is T2-WI of diagnostic quality?  
- Yes  
- No

#### DWI

**Technical parameters**
- Axial plane matching T2-WI
- Adequate field of view
- Adequate in plane resolution
- Adequate slice thickness
- Multibar (≥ 3) values acquired
- High b-value (synthesised or sequenced)

**Visual assessment**
- Adequate ADC map
- Absence of artifacts (e.g. metal artefacts)

Is DWI of diagnostic quality?  
- Yes  
- No

#### DCE

**Technical parameters**
- Axial plane matching T2-WI
- Adequate field of view
- Adequate in plane resolution
- Adequate slice thickness
- Pre-contrast T1-WI available
- Fat suppression/subtraction
- Adequate temporal resolution (≤ 10 sec)
- Adequate total observation rate (≥ 20min)

**Visual assessment**
- Capillary vessels clearly delineated
- Vessels in the Alcock's canal clearly delineated
- Absence of artifacts (e.g. movement)

Is DCE of diagnostic quality?  
- Yes  
- No

#### PI-QUAL score:
- 1
- 2
- 3
- 4
- 5

**Comments:**

**Date:**

**Reporting Radiologist:**

**Signed:**


**Figure 6-1 PIQUAL Scoring system**
6.2 METHODS

6.2.1 PARTICIPANTS

The participants in this study were from the INNOVATE trial as described previously. This report includes a random sample of 57 taken from the cohort of 303. This sample size was based on reader availability and the time taken to read each dataset. Randomisation was performed on Microsoft Excel (version 16.49, Microsoft, 2021).

The study cohort was divided into two groups: A and B. For cohort A, datasets were compiled with T2W coronal, T2W axial, high b value (b1400 or b2000), dynamic contrast imaging and FIC map. In Cohort B, datasets were compiled with ADC maps instead of FIC maps (Figure 6.2).

An equal number of datasets from Cohort A and Cohort B were taken for each reading session. Each reading session was typically comprised of 20 datasets, of which 10 were FIC maps and 10 were ADC maps. This was to ensure that readers had equal exposure to the two maps rather than imbalance which may affect results.

Figure 6-2 Study Design
6.2.2 READER TRAINING

To familiarise the readers with VERDICT FIC maps, both readers were given a structured training session consisting of 25 cases before reading for the study began. The 25 cases were taken from a training cohort of men recruited in the INNOVATE study who were undergoing active surveillance. These participants were not included in the subsequent study. The imaging data set included T2W coronal, T2W axial, high b value (b1400 or b2000), dynamic contrast imaging, ADC and FIC maps. Readers were given examples of mpMRI scans scored Likert 2, Likert 3, Likert 4, and Likert 5 by a group of uro-radiologists at UCLH. Both ADC and FIC maps were assessed together by readers to familiarise themselves with the appearances of biopsy-proven prostate cancer on both image types. Biopsy and clinical outcomes were available for the readers.

6.2.3 IMAGE QUALITY

Readers assessed and scored each image type individually in a locked sequence starting with T2W axial and coronal, followed by either ADC or FIC, high b value and dynamic contrast enhancement (Figure 6-3). First, image quality was assessed using the PIQUAL and subjective Likert scoring systems. The PIQUAL system uses a checklist of technical parameters and visual inspection to determine whether each image type is diagnostic or non-diagnostic. The overall PIQUAL score depends on how many image types are rated as diagnostic in a dataset for a participant. Although PIQUAL does not specifically include the assessment of VERDICT maps, it does have specific criteria for diffusion-weighted imaging. This criteria were applied by readers to assess VERDICT maps. The subjective Likert score for image quality uses a 5-point scale as follows:

1: very poor quality, considered non-diagnostic (artefacts on all slices, scans uninterpretable)
2: poor quality with some impairment of diagnostic quality (substantial artefacts, but still interpretable),

3: satisfactory quality without impairment of diagnostic quality (some artefacts present),

4: good quality (hardly any artefacts),

5: excellent quality (no artefacts present).

The main artefacts found on both ADC and FIC maps are susceptibility artefacts from rectal gas as shown in Figure 6-4.

**Figure 6-3** Locked Sequential Read scheme

FIC = fractional intracellular volume, ADC = apparent diffusion coefficient, DCE = dynamic contrast enhancement. The highlighted ‘Likert score’ and ‘PIRADS 2.1 score’ are the scores that were assessed for the primary and secondary outcomes.
Figure 6-4 ADC maps (left) and FIC maps (right) from 5 participants.

The red letter is the Likert rating given by both readers, for example, '5' was rated Likert 5 (excellent quality), '4' (good quality), '3' (satisfactory quality), '2' (poor quality), '1' (very poor quality). Red arrows show susceptibility artefact from rectal gas, which are worse for lower quality scores.
Both readers had more than five years of reporting prostate mpMRI. Each image type was scored in a locked sequence as described above (T2W then FIC or ADC, then high b value, followed by DCE), using the subjective Likert scoring used at our institution and PIRADS 2.1 by two readers. The prostate was divided into quadrants at the base, midgland and apex as shown in Figure 6-6. The quadrant scoring was based on Likert and used for each image type. The locked sequential scoring allows assessment of the added value of sequences\[97\]. A PIRADS 2.1 score was given to the entire prostate in addition to an overall Likert score. Readers were instructed to draw any focal lesions or diffuse changes on a schematic of the prostate and score individual lesions.

### 6.2.5 TIME TAKEN

The time taken by readers to score the first 15 cases and the last 15 cases was recorded, for instance from case 1 to case 15 and case 43 to case 57. In particular, the time taken to read the FIC or ADC map was also recorded. The rationale for recording the first and last 15 cases was to capture the learning curve for reading FIC maps. Readers were less familiar with FIC maps compared to ADC maps and we hypothesised that readers would take less time for the last 15 reads compared to the first 15. The reading session was recorded on Microsoft Teams (Version 1.4, 2021) which allowed for accurate timekeeping.

### 6.2.6 REFERENCE STANDARD

The reference standard comprised of biopsy for men who underwent transperineal or transrectal targeted biopsy was histology (Figure 6-4). Clinically significant cancer was defined as any biopsy core containing Gleason 3+4 disease. For the men who did not undergo biopsy, a clinical mpMRI score of 2 or below was
assigned as negative. This is based on the high negative predictive value (96.1\%) of mpMRI for clinically significant cancer (Gleason grade group $\geq 3$) [98].

For those men who had a score of Likert 3 or above and who did not undergo biopsy, the reference standard was based on follow up as per clinical care (Figure 6-5). These men are followed up with repeat PSA at 6 monthly intervals and if the PSA rises, mpMRI is repeated.

To derive a false positive, true positive, false negative and false positive rate, Likert 3 or above scores from readers were considered as positive. The rationale for this is based on similar practice in multiple previous studies and clinical practice [23,63,86,98]. Men who have Likert 1 or 2 scores are not offered biopsy.

A sub-group analysis was also carried out for a reference standard only including men who had a biopsy or were scored Likert 2 on mpMRI or subsequent follow-up imaging (n=42/57).
6.2.7 **Statistical Analysis**

Image quality was compared for FIC and ADC maps using the non-parametric paired test called the Wilcoxon Signed Ranks test. Overall PIQUAL scores for datasets with FIC were compared to datasets with ADC using the same test. To account for differences in image quality due to different field strength of the two scanners, a subgroup analysis of image quality was carried out for ADC maps acquired at 1.5T and at 3T.

Inter-observer agreement was assessed using Cohen’s kappa (K) and Gwet’s agreement co-efficient (AC). The reason for using both is that Cohen’s kappa can be
unexpectedly low even in the presence of high agreement due to reported paradoxes resulting from methodology [99].

Inter-reader agreement was analysed for three levels based on how Likert/PIRADS scores are used clinically. For the first subgroup, scores of 1 or 2 were defined as negative, scores of 4 or 5 were defined as positive and a score of 3 was assigned as indeterminate. The values were interpreted according to Landis and Koch [100], with the following definitions: $K < 0$ indicates no agreement, 0–0.20; slight, 0.21–0.40; fair, 0.41–0.60; moderate, 0.61–0.80; substantial and 0.81–1 as almost perfect. Inter-observer variability in image quality assessment was also compared using Cohen’s kappa.

Analyses were carried out at two points for reader scores in the locked sequential scheme (Figure 6-3). The first was when readers scored either the T2W and FIC or T2W and ADC images. The highest Likert score was taken as the overall score. These scores are described as ‘individual FIC’ or ‘individual ADC’ reads. The second was the overall Likert score and PIRADS 2.1 score for the prostate at the end of the locked sequential read. The scoring proforma is provided in Figure 6-5.

For this analysis, Likert and PIRADS scores equal to or more than 3 were defined as positive and scores of 1 or 2 were defined as negative. This methodology has been used in several studies and reflects the clinical management [11,91,98]. For the primary objective, the reference standard was applied to derive the number of true positives, false positives, true negatives, and false negatives.

Differences between FIC and ADC reads for individual image types and the whole dataset were assessed using the McNemar test which is a non-parametric test for paired data.
Figure 6-6 Reporting pro-forma for the Qualitative study
6.3 RESULTS

6.3.1 PARTICIPANTS

The demographics of the INNOVATE cohort have been mentioned previously. The random sample of 57 men had a median age of 64 (range 44–78) and a median PSA of 5.98 ng/ml (range 0.83–37.4).

6.3.2 IMAGE QUALITY

Image quality was rated using two different scoring systems by two readers. The frequencies of scores are charted in Figures 6-6 and 6-7. For the subjective Likert scoring, the mean FIC quality score for reader 1 was 3.5 (IQR 1) and 3.4 (IQR 2) for reader 2. The mean score for ADC for reader 1 was 3.2 (IQR 1) and 3.5 (IQR 2) for reader 2. There was a statistical difference between FIC quality and ADC quality for reader 1 (p=0.021) but not for reader 2 (p=0.663).

For PIQUAL scoring, the mean quality rating for datasets that had FIC was 3.5 for reader 1 (IQR 1) and 4.1 (IQR 1) for reader 2. For ADC, the mean PIQUAL score was 3.4 (IQR 1) for reader 1 and 4 (IQR 2) for reader 2. There was no statistical difference between FIC and ADC PIQUAL scores (p=0.240 and p=0.614).

There was a similar number of cases that were rated as poor quality or undiagnostic for FIC and ADC by both readers for Likert scoring. For instance, the first reader rated 9 FIC cases as 2/5 for image quality and 9 ADC cases as 2/5 on the Likert scale. In addition, the first reader rated 4 cases as 1/5 for ADC maps. The second reader rated 14 FIC and 14 ADC cases as 2/5 on the Likert quality scale. One FIC and ADC case was rated 1/5 by the second reader.
6.3.2.1 1.5 T AND 3T subgroup

For participants who had ADC maps derived from 1.5 T data (n=18), the mean ADC quality rating for reader 1 was 2.9 and for reader 2 was 3.3. The mean FIC quality score was 3.1 for reader 1 and 3.0 for reader 2 in this subgroup. There was no difference between ADC and FIC quality scores in this subgroup for reader 1 (p=0.542) and reader 2 (p=0.263). The mean PIQUAL score for ADC datasets was 3.1 for reader 1 and 3.8 for reader 2. The mean PIQUAL score for FIC datasets for reader 1 was 3.2 and 4.1 for reader 2. There was no difference between PIQUAL scores for FIC compared to ADC for reader 1 (0.579) and reader 2 (p=0.083).

For participants who had ADC maps derived from 3T data (n=39), the mean ADC quality rating for reader 1 was 3.4 and for reader 2 was 3.5. The mean FIC quality score was 3.7 for reader 1 and 3.6 for reader 2. For reader 1, FIC quality score was higher than ADC (p=0.022). There was no difference between ADC and FIC quality scores in this subgroup for reader 2 (p=0.72). The mean PIQUAL score for ADC datasets was 3.5 for reader 1 and 4.1 for reader 2. The mean PIQUAL score for FIC datasets for reader 1 was 3.6 and 4.0 for reader 2. There was no difference between PIQUAL scores for FIC compared to ADC for reader 1 (p=0.324) and reader 2 (p=0.549).

The mean ADC image quality for ADC maps derived from 1.5T was 2.9 for reader 1 and 3.0 for reader 2. The mean ADC image quality for ADC maps derived from 3T was 3.4 for reader 1 and 3.5 for reader 2. There was no difference in image quality between scanners for ADC maps as rated by reader 1 (p=0.131) or reader 2 (0.405).
Figure 6-7 Likert Quality Scores for both readers; FIC on left, ADC on right

Figure 6-8 PIQUAL Quality scores for both readers
6.3.3 INTER-READER AGREEMENT

Interobserver agreement was measured between readers for individual maps and overall FIC and ADC datasets.

The interobserver agreement was ‘fair’ for both individual FIC and ADC maps; with FIC having a slightly higher kappa of 0.371 compared to 0.323 for ADC. The Gwet’s agreement coefficient (AC) was comparable for both FIC and ADC maps at 0.369 and 0.399 respectively. The percentage agreement for both was identical at 57.9%. The agreement was highest for negative scores of Likert 1 or 2 at 83%.

Interobserver agreement for FIC increased when readers evaluated the whole dataset and decreased for ADC datasets. For FIC datasets, the agreement was moderate with a kappa value of 0.548 and for ADC it was fair at 0.285. Similarly, the AC value for FIC was 0.579 and 0.368 for ADC. The percentage agreement for ADC was 56% and was higher for FIC at 72%.

Using the PIRADS scoring system, the interobserver agreement for FIC was comparable to Likert scoring at 0.556 (moderate agreement). However, for ADC, the agreement was lower at 0.232. Similarly, the AC value for FIC datasets was 0.597 (approaching substantial agreement) and fair for ADC at 0.318. The percentage agreement for ADC was 53% and for FIC was 72%.

6.3.4 READING TIME

For the first 15 cases, the mean reading time for FIC maps for reader 1 was 33 seconds (range 20-60s) and for reader 2 was 120 seconds (range 50-240s). For ADC maps, for reader 1 the mean reading time was 46 seconds (range 30-120s) and 100 seconds (range 40-250s) for reader 2. The difference between ADC and FIC for both readers was not statistically significant (p=0.066 and p=0.362).
The mean time to read the whole FIC dataset for reader 1 was 299 seconds/4 minutes 59 seconds (range 180-480 s) and for reader 2 was 543 seconds/9 minutes 3 seconds (range 250-870 s). For ADC datasets, the mean reading time for reader 1 was 278 seconds/4 minutes 38 seconds (range 230-370 s) and for reader 2 was 411 seconds/6 minutes 51 seconds (range 300-540 s). This difference was not statistically significant (p=0.191).

For the last 15 cases, the average time for reader 1 for FIC maps was 23 seconds (range 20-30 s) and for reader 2 was 102 seconds (50-120 s). For ADC maps, this was 42 seconds (range 30-55) for reader 1 and 86 seconds for reader 2 (range 30-200 s). Although readers were slightly faster for the last 15 cases for FIC (reader 1: p=0.088 and reader 2: p=0.462)) and had comparable times for ADC maps (p=0.721 and p=0.773), the differences were not statistically significant.

For overall FIC datasets, the mean time for reader 1 was faster at 240 seconds/4 minutes (range 180-300 s) and for reader 2, it was 393 seconds/6 minutes 33 seconds (range 250-600s) for the last 15 cases. For ADC datasets, reader 1’s mean reading time was 244 seconds/4 minutes 4 seconds (range 180-300 s) and for reader 2 it was 446 seconds/7 minutes 26 seconds (240-600 s). Although readers were faster reading FIC datasets (reader 1 p=0.09, reader 2 p=0.127) for the last 15 cases and had comparable times for ADC datasets (p=0.232 and p=0.426), the difference did not reach statistical significance.
6.3.5 Reference Standard

A total of 32 (56%) men underwent a biopsy out of 57 (Table 6-1). 15 out of 32 men had clinically significant cancer; with 12 biopsies showing Gleason 3+4 disease and 3 biopsies showing Gleason 4+3 disease. 3 out of 32 had clinically insignificant cancer (Gleason 3+3) and 14 men had a negative biopsy.

In those men who did not undergo biopsy, 9/25 had a Likert score of 2 and were discharged from hospital follow up. These were defined as negative.

Three men had a Likert score of 3 and a low presenting PSA density of 0.055, 0.057 and 0.074. They were considered to have a low probability of clinically significant cancer and were discharged from hospital follow up with no re-referral after 3-4 years follow-up. These cases were defined as negative.

The remaining 12 men had a score of Likert 3 and underwent PSA/imaging surveillance. The average follow-up was 13 months (range 6-36 months). For 9 out of 12 men, the PSA level decreased. These were assigned as negative.

The remaining 3 men with Likert 3 had stable PSA values on PSA surveillance. One of these men had repeat imaging which was scored as Likert 2. One of these men had a stable PSA and stable imaging after 2 years. The other man had a stable PSA after 1 year. None of these men was offered a biopsy in their follow up period (6-36 months) and were considered as low risk to have csPCa. There were assigned as negative.
### Table 6-1 Application of reference standard to study cohort

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Biopsy</td>
<td>15</td>
</tr>
<tr>
<td>Negative Biopsy or Gleason 3+3</td>
<td>18</td>
</tr>
<tr>
<td>Negative mpMRI (Likert 2)</td>
<td>9</td>
</tr>
<tr>
<td>Not biopsied, low PSAD/no progression in PSA or Imaging</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>57</strong></td>
</tr>
</tbody>
</table>

#### 6.3.6 DIAGNOSTIC ACCURACY

After the application of the reference standard, the number of true positives, false negatives, false positives, and true negatives for individual FIC vs ADC and multiparametric reads (including all image types) is given in Table 6-2.

The mean number of false positives for individual FIC reads (24/57) was lower compared to ADC reads (33/57). This difference was statistically significant (p=0.041). The mean number of true negatives was higher at 20 for FIC compared to 10 for ADC. This difference was statistically significant (p=0.041). The mean number of true positives for FIC reads was the same as ADC reads; 13 for both. The difference was not statistically significant (p>0.99). The mean number of false negatives for FIC reads was the same as ADC reads at 1/57. The difference was not statistically significant (p>0.99).
Table 6-2 Application of overall reference standard to Reader scores (+) denotes entire dataset including T2W, FIC or ADC, high b-value image and DCE images.

TP = True Positives, FN = False Negatives, FP = False Positives, TN = True Negatives, Spec = specificity, Sens = sensitivity, PPV = positive predictive value, NPV = negative predictive value.

<table>
<thead>
<tr>
<th></th>
<th>Reader 1</th>
<th>Reader 2 (LIKERT)</th>
<th>Reader 1 (LIKERT)</th>
<th>Reader 2 (PIRADS)</th>
<th>Reader 1 (PIRADS)</th>
<th>Reader 2 (PIRADS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIC</td>
<td>ADC</td>
<td>FIC + ADC + FIC + ADC +</td>
<td>FIC + ADC + FIC + ADC +</td>
<td>FIC + ADC + FIC + ADC +</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>FN</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FP</td>
<td>20</td>
<td>28</td>
<td>27</td>
<td>38</td>
<td>31</td>
<td>35</td>
</tr>
<tr>
<td>TN</td>
<td>23</td>
<td>15</td>
<td>16</td>
<td>5</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Spec</td>
<td>53%</td>
<td>35%</td>
<td>37%</td>
<td>12%</td>
<td>28%</td>
<td>19%</td>
</tr>
<tr>
<td>Sens</td>
<td>93%</td>
<td>86%</td>
<td>93%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>PPV</td>
<td>39%</td>
<td>30%</td>
<td>33%</td>
<td>27%</td>
<td>31%</td>
<td>29%</td>
</tr>
<tr>
<td>NPV</td>
<td>96%</td>
<td>88%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The overall Likert score for the multiparametric dataset which included all the image types with either FIC or ADC showed a similar trend. The mean number of false positives was also lower for datasets that included FIC (31/57) compared to ADC (34/57) for Likert scoring. The difference was not statistically significant (p=0.264). The number of true negatives was higher for FIC (12) compared to ADC (9) but not significantly different (p=0.264). The number of false negatives was 0 for both FIC and ADC reads. The true positives were the same for both image types at 15/57.

For PIRADS scoring, the number of false positives was lower for FIC datasets (16/57) compared to ADC (28/57). This difference however was not statistically significant (p=0.189). The number of true positives was the same for FIC (13) and ADC (13) with no statistically significant difference (p>0.99). The number of true
negatives was higher for FIC (27) compared to ADC (23); however, this was not statistically significant (p=0.189).

The overall specificity and positive predictive value were consistently higher for FIC compared to ADC. The sensitivity and negative predictive value were similar for both.

6.3.7 SUB-GROUP ANALYSIS

A similar analysis to above was carried out for those men who had a biopsy or were scored Likert 2 on initial or subsequent MRI.

The mean number of false positives for individual FIC reads (16/57) was lower compared to ADC reads (19/57). This difference was not statistically significant (p=0.065). The mean number of true negatives was higher at 12 for FIC compared to 8 for ADC. This difference was not statistically significant (p=0.065). The mean number of true positives for FIC reads was the same as ADC reads; 13 for both. The difference was not statistically significant (p>0.99). The mean number of false negatives for FIC reads was the same as ADC reads at 1/57. The difference was not statistically significant (p>0.99).

For the multiparametric dataset which included all the image types with either FIC or ADC, the metrics were comparable. The mean number of false positives was 20 for both FIC and ADC datasets. The difference was not statistically significant (p=0.424). The mean number of true negatives was also the same at 8 for each and not significantly different (p=0.424). The number of false negatives was 0 for both FIC and ADC reads. The true positives were the same for both image types at 15/57.
For PIRADS scoring, the number of false positives for FIC datasets were similar at \((12/42)\) compared to ADC \((13/42)\). This difference was not statistically significant \((p=0.267)\). The number of true positives was similar for FIC \((13)\) and ADC \((12)\) with no statistically significant difference \((p>0.99)\). The number of true negatives was higher for FIC \((16)\) compared to ADC \((14)\); however, this was not statistically significant \((p=0.267)\).

Table 6-3 Application of sub-group reference standard to Reader scores (+) denotes entire dataset including T2W, FIC or ADC, high b-value image and DCE images.

TP = True Positives, FN = False Negatives, FP = False Positives, TN = True Negatives, Spec = specificity, Sens = sensitivity, PPV = positive predictive value, NPV = negative predictive value.

<table>
<thead>
<tr>
<th>Reader 1</th>
<th>Reader 2</th>
<th>Reader 1 (LIKERT)</th>
<th>Reader 2 (LIKERT)</th>
<th>Reader 1 (PIRADS)</th>
<th>Reader 2 (PIRADS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIC</td>
<td>ADC</td>
<td>FIC</td>
<td>ADC</td>
<td>FIC (+)</td>
<td>ADC (+)</td>
</tr>
<tr>
<td>TP</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>FN</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FP</td>
<td>14</td>
<td>17</td>
<td>18</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>TN</td>
<td>14</td>
<td>11</td>
<td>10</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Spec</td>
<td>50%</td>
<td>39%</td>
<td>36%</td>
<td>19%</td>
<td>25%</td>
</tr>
<tr>
<td>Sens</td>
<td>93%</td>
<td>86%</td>
<td>93%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>PPV</td>
<td>48%</td>
<td>41%</td>
<td>42%</td>
<td>40%</td>
<td>41%</td>
</tr>
<tr>
<td>NPV</td>
<td>93%</td>
<td>85%</td>
<td>91%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

6.3.8 **LOCKED SEQUENTIAL ANALYSIS**

To investigate the difference in false positives between individual FIC or ADC map scores to overall dataset reads, the locked sequential scores were examined on a case-by-case basis for each reader. In particular, where the overall dataset Likert
score was different to the individual FIC or ADC score, I examined which image type influenced the final Likert score.

The number of men who had an FIC image score of 2 but the overall dataset Likert score was different was 11 for the first reader. In these men, the high b value image was scored differently in 5/11 men (4 x Likert 3, 1 x Likert 4) and DCE in 11/11 men. DCE was scored Likert 3 in 10/11 and Likert 4 in 1/11. The overall Likert score was Likert 3 for 10/11 and one man was scored Likert 4/5.

For those men who had an FIC score of 2 from the second reader but the overall dataset Likert score was different (12 men), the high b value image was scored differently in 5/12 men (2 x Likert 3, 3 x Likert 4) and DCE in 10/12 men. DCE was scored Likert 3 in all these cases. The overall Likert score for these men was Likert 3 for 10/12 and two men were scored Likert 4.

For those men who had an ADC score of 2 from the first reader but the overall dataset score was different (10 men), the high b value image was scored differently in 3/10 men (2 x Likert 3, 1 x Likert 4) and DCE in all 9/10 men (8 x Likert 3, 1 x Likert 4). The overall Likert score for these men was Likert 3 for 10/12 and two men were scored Likert 4.

For those men who had an ADC score of 2 from the second reader but the overall dataset score was different (3 men), the high b value image was scored differently in 2 men (2 x Likert 3) and DCE in all three men (2 x Likert 3, 1 x Likert 4). The overall Likert score for these men was Likert 3 for 2/3 and one man were scored Likert 4.
6.4 DISCUSSION

The major findings of this study are that the image quality of FIC was comparable to ADC as rated by two experienced radiologists. Inter-reader agreement using two scoring systems was modest and similar for both FIC and ADC regardless of whether the Likert and PIRADS scoring system is used. Furthermore, preliminary analysis of the cohort shows that assessment of FIC maps could reduce the number of false-positive results compared to ADC when read individually. There was also a reduction in false positives when readers scored the entire dataset with FIC compared to ADC, however, this was not statistically significant.

The image quality scores for FIC maps were comparable to ADC maps for both scoring systems used in this study. Although one reader rated FIC maps as better quality compared to ADC using Likert scoring. This is expected as both acquisitions use echo planar imaging which is susceptible to susceptibility artefacts from rectal gas or peristalsis.

Inter-observer agreement was fair for FIC and ADC maps when read individually and increased to moderate/substantial agreement when the entire dataset was read for FIC but not for ADC. In this analysis, it was noted that Gwet’s agreement coefficient was generally higher than the kappa value. This is likely because the percentage agreement was high in the comparisons made and kappa can be paradoxically low in such conditions[99]. The values of the agreement metrics are similar to the level of agreement previously published for multiparametric MRI[11,101]. The modest agreement seen in this study and previously published studies is likely due to the MRI conspicuity of prostate cancer. Prostate cancer is a heterogeneous disease and can be multifocal. As discussed previously, benign conditions can also mask the appearance of prostate cancer. Readers have different thresholds for calling something indeterminate and varying levels of confidence in
each image type which leads to variability. The reading experience of both readers was similar and therefore this is unlikely to explain the variability. Interestingly, the inter-reader agreement was similar for PIRADS which is a more prescriptive and structured reporting system. This finding supports the notion that it is not the reporting system that readers use but the assessment of a heterogeneous disease that leads to variability.

Overall readers took a similar amount of time to assess FIC maps compared to ADC maps. This was also the case for reading the whole dataset which either contained FIC or ADC maps. There was some evidence of a learning curve for FIC as both readers were faster to read the last 15 cases compared to the first 15. Readers took a similar time to read ADC maps for the first 15 and last 15, which is not surprising as both readers had over 5 years of experience reading ADC maps.

The reduction in false positives from FIC maps compared to ADC is likely due to differences in sequence design and model fitting. VERDICT MRI is a more complex model designed specifically for the prostate with three compartments whereas ADC is a simpler mathematical model which is known to have limitations in fitting at low and higher b values. The acquisition for VERDICT uses higher b-values which are accounted for in the mathematical model to increase biological specificity. Therefore, false positives caused by benign pathologies such as atrophy and inflammation may have a lesser impact on VERDICT MRI. The VERDICT model may be able to mitigate these effects by separating these diffusion environments.

The number of false positives was lower for readers when assessing individual FIC maps compared to the whole dataset including T2W, high B value and DCE imaging. This is likely due to either indeterminate appearances on other sequences included in the multiparametric dataset (particularly DCE) or a lack of familiarity or trust of FIC maps. Both readers had more than 5 years of experience reporting multiparametric MRI and no prior experience of assessing FIC maps. Therefore, it
is plausible that confidence in a new technique would be limited, especially when blinded to the patient outcome or PSA.

The limitations of this analysis are that for men who did not undergo biopsy there was no definitive histological reference standard. However, as the negative predictive value of mpMRI is high, the assumption that men who have a negative mpMRI have no significant cancer is not unreasonable. The relatively small number of patients included in this analysis is also a limitation but taking a random sample from the cohort should be representative of the larger cohort. Readers were blinded to PSA which is not the case for clinical Likert scoring but is recommended for PIRADS scoring. Therefore, the recorded Likert scores may not be representative of clinical practice. Finally, FIC maps were derived from VERDICT acquisition on a 3T scanner whereas some ADC maps (n=18) were derived from 1.5T and this could be a confounding factor in assessing image quality. However, subgroup analysis of participants who had both datasets derived from 3T, did not show any differences in image quality between FIC and ADC.

**Conclusion**

Image quality and inter-observer variability were comparable for FIC maps compared to ADC maps. This interim analysis of the INNOVATE cohort also shows qualitative assessment could reduce false positives from mpMRI but perhaps not to the degree that quantitative analysis can.
7 The effect of hormonal therapy and radiotherapy on VERDICT and multiparametric MRI parameters

Author declaration

All the work in this chapter was conceived and written by me, under the supervision of Dr Anita Mitra and Professor Shonit Punwani. VERDICT maps were produced via an automatic fitting process using a cloud-based platform called XNAT (maintained at UCL by Baris Kanber), which integrates the VERDICT code of Dr Eleftheria Panagiotaki.

7.1 INTRODUCTION

External beam radiotherapy (EBRT) is an established treatment option in the management of organ-confined and locally advanced prostate cancer. Despite improvements in the delivery of radiotherapy, a proportion of patients go on to have recurrence after therapy. The rate of positive biopsies after EBRT reported in the literature ranges from as low as 4.2% and as high as 67%.

Currently, the follow up of patients who have undergone radiotherapy is with serum PSA measurement. The lowest PSA value after treatment is taken as the nadir value and if the PSA rises by 2ng/ml above this level, this is deemed Biochemical failure (BCF) by the Phoenix criteria [102]. BCF after EBRT is the most established predictor of PCa recurrence. Despite its obvious benefits, there are many limitations. Patients can have positive biopsies for recurrence before PSA rises [103]. Another drawback of BCF is that it does not discriminate between locally recurrent disease and metastasis. PSA as discussed previously is increased in benign conditions and can cause false positives.

Increasingly for the assessment of local tumour control, imaging and biopsy are used [104]. Both multiparametric magnetic resonance imaging (mpMRI) and prostate-specific membrane antigen-positron emission tomography (PSMA-PET)
have shown promise in detecting residual or recurrent cancer after EBRT but can be difficult to interpret due to treatment-related changes in the prostate and spatial resolution \[^{104-106}\]. Therefore, to confirm recurrence or residual disease, a biopsy is usually performed.

Biopsy has also been investigated as a stand-alone technique to evaluate local PCa recurrence after EBRT, albeit much less frequently \[^{107}\]. Post-EBRT biopsies taken less than two years after EBRT treatment are not reliable and can still be challenging to interpret, for instance when cancer and treatment effect is leading to ‘indeterminate’ reports alongside distinctly positive and negative \[^{107}\].

Due to the varying positive biopsy rates and the various factors affecting local disease control, I carried out a meta-analysis to assess how many men have residual disease after treatment and whether the presence of the disease has an impact on disease-free survival, metastatic disease, and overall mortality (Appendix).

The meta-analysis showed the weighted-average positive biopsy rate across all 22 studies was 32\% (95\%-CI: 25-39\%, n=3017). In studies where post-treatment biopsy as part of the study protocol, the rate was 35\% (95\%-CI: 21-38\%, n=2450). In the subgroup of studies that conformed to the 2020 NCCN radiotherapy guidelines, this rate was 22\% (95\% CI: 19-11, n=832). Patients with positive biopsy had a 10-fold higher odds of developing BCF \((\text{OR} \ 10.3, 95\%-\text{CI}: \ 3.7-28.7, p<0.00001)\), 3-fold higher odds of developing distant metastasis \((\text{OR} \ 3.1, 95\%-\text{CI}: \ 2.1-4.7, p<0.00001)\) and 5-fold higher odds of dying from their PCa \((\text{OR} \ 5.1, 95\%-\text{CI}: \ 2.6-10, p<0.00001)\) \[^{103}\].

Recurrence or residual disease after therapy is associated with poor oncological outcomes. Therefore, there is a clinical need to detect recurrence as early as possible. It is also apparent from the literature that PSA monitoring can miss men with the disease after therapy.
Apparent Diffusion Coefficient (ADC) values after radiotherapy have been shown to change earlier than PSA\textsuperscript{[103]} and therefore could help in monitoring response to therapy. This change in diffusion could be due to cell lysis or a change in vasculature after therapy.

VERDICT MRI could have a role in monitoring therapy if diffusion is significantly altered in a treated prostate. In the previous chapters, it has been demonstrated that it may have greater biological specificity compared to ADC. Serial changes in VERDICT parameters may better reflect histological changes and allow for the detection of recurrence.

The objective of this study is to evaluate the changes in VERDICT and mpMRI quantitative parameters in localised or locally advanced prostate cancer after androgen deprivation therapy (ADT). Patients were scanned before androgen deprivation therapy (ADT) started, 3 weeks before the start of radiotherapy, again in week 6 of radiotherapy and once more 6 months after the end of radiotherapy.

7.1.1 AIMS

i) To determine whether VERDICT and mpMRI parameters change after ADT

ii) To determine whether VERDICT and mpMRI parameters change after EBRT

7.1.2 NULL HYPOTHESES

iv) There is no change in VERDICT and mpMRI parameters change after ADT

v) There is no VERDICT and mpMRI parameters change after EBRT
7.2 METHODS

7.2.1 PARTICIPANTS

The inclusion criteria will be men with a biopsy confirmed prostate cancer who opt for ADT and EBRT.

The exclusion criteria were:

i. Treatment within the previous 6 months with any form of hormones (including 5-alpha reductase inhibitors)
ii. Evidence of metastatic disease
iii. Prior local intervention to the prostate
iv. Unable to give informed consent
v. Any prosthesis (including hip replacements) which could cause artefacts degrading the quality of the imaging
vi. Contraindication to gadolinium contrast agent
vii. Unable to tolerate an MRI

7.2.2 STUDY DESIGN

The study design will follow the standard of care for patients undergoing ADT and EBRT for prostate cancer with the addition of a research VERDICT MRI as outlined in Figure 9-1. The sequence parameters were identical to those used in the INNOVATE study.
Figure 7-1 Study Schedule

7.2.3 IMAGE ANALYSIS

The biopsy positive tumour was selected by reviewing the biopsy report, MRI images and pictorial report with biopsy targets. A Region of Interest (ROI) was drawn on a single slice containing the largest portion of the tumour on ADC and VERDICT-derived maps (Figure 9-2).
Tumour-free regions of the peripheral zone and transition zone were also selected by reviewing the aforementioned information. The same region was used for subsequent ROIs unless there was artefact. In the presence of artefact, an adjacent tumour free region was selected.

PSA density was calculated by dividing the serial PSA measurements by MRI-calculated prostate volume using the ellipsoid method. Gland Volume was measured at each time point.

Shapiro-Wilk test was used to test the data for normality. Non-parametric tests were used to compare differences between parameters for each time point. The time points were coded as:

1 – Before ADT

2- After ADT, 3 weeks before EBRT

3- 6 weeks during EBRT

4- 6 months after EBRT

Differences in parameters between time points were compared using a non-parametric test for related samples called Friedman’s two-way analysis of variance by ranks. A Bonferroni correction for multiple tests was applied for significance testing. Statistical analyses were performed using SPSS v.26 (IBM Corp., Armonk, NY, USA). The p-value threshold for significance was set at p < 0.05.
**Figure 7-2** MRI Images of a 78-year-old man with a Gleason 3+4 tumour in the anterior left prostate.

Axial images through the midgland from top to bottom: T2W, Apparent Diffusion Coefficient (ADC), fractional intracellular volume (FIC) map, fractional extracellular extravascular space (FEES), fractional vascular fraction (FVASC)
7.3 RESULTS

7.3.1 PARTICIPANTS

5 men completed the study protocol. The median age was 78 (range 75-79). All men had intermediate (1 man had 3+4, 3 men had 4+3) or high-risk disease (3+5) on biopsy. The average number of days between starting hormones and having the 2nd MRI was 216 days/7 months 4 days (IQR 90). The average number of days between starting hormones and starting EBRT was 242 days/7 months, 30 days (IQR 7.5). The interval between the starting EBRT and the 3rd MRI was on average 41 days (IQR 7). The average interval between starting EBRT and the 4th MRI was 237 days/7 months 25 days (IQR 35).

7.3.2 PSA AND PSA DENSITY

Before treatment, the median PSA for the cohort was 9.29 ng/ml (range 5.8-26.9) and the mean PSA: 14.02 ng/ml. There was a 97% decrease in mean PSA after ADT (from 14.02 to 0.39 ng/ml, p=0.518). There was a further 86% reduction in mean PSA after 6 weeks of radiotherapy (0.39 to 0.08 ng/ml, p>0.99) and a smaller 26% reduction (0.08 to 0.03 ng/ml, p>0.99), 6 months after radiotherapy. There was a statistically significant difference between PSA before any treatment (time point 1) and 6 months after EBRT (time point 4); p=0.009 but not between other time points.

There was a similar trend with PSAD. The median PSAD before any therapy was 1.416 ng/ml/ml and mean PSAD 0.404 ng/ml/ml. This decreased by 97% to 0.013 after hormones (p>0.99) and 0.004 after 6 weeks of EBRT (p=0.850). The final mean PSAD after at least 6 months of EBRT was similar at 0.001 (p>0.99). Both trends are demonstrated in Figure 9-3. There was a statistically significant difference between PSA before any treatment (time point 1) and 6 months after EBRT (time point 4); p=0.02.
Before treatment, the mean lesion ADC was $0.957 \times 10^{-3}$ mm²/s (range 0.762-1.111), and median ADC was $1.00 \times 10^{-3}$ mm²/s. After ADT, the mean lesion ADC increased by 6% to $1.01 \times 10^{-3}$ mm²/s. At time point 3, after 6 weeks of radiotherapy, the mean ADC increased by 14% to 1.159. After 6 months, ADC

---

**Figure 7-3** Trend of PSA and PSA density (PSAD) for 5 patients.

Time points are defined as: (1) – Before androgen deprivation therapy (ADT) (2) - After ADT, 3 weeks before external beam radiotherapy (EBRT) (3) - 6 weeks during EBRT (4) - 6 months after EBRT

---

### 7.3.3 *Apparent Diffusion Coefficient*

Before treatment, the mean lesion ADC was $0.957 \times 10^{-3}$ mm²/s (range 0.762-1.111), and median ADC was $1.00 \times 10^{-3}$ mm²/s. After ADT, the mean lesion ADC increased by 6% to $1.01 \times 10^{-3}$ mm²/s. At time point 3, after 6 weeks of radiotherapy, the mean ADC increased by 14% to 1.159. After 6 months, ADC
remained stable, increasing by 1% to $1.234 \times 10^{-3}$ mm$^2$/s. There were no significant differences between ADC at different time points ($p=0.145$).

For tumour-free regions of the peripheral zone, the mean ADC before therapy was $1.659 \times 10^{-3}$ mm$^2$/s. The mean ADC after hormones was $1.398 \times 10^{-3}$ mm$^2$/s, 1.505 after 6 weeks of EBRT and $1.391 \times 10^{-3}$ mm$^2$/s after 6 months of EBRT. There was no statistically significant difference between these time points ($p=0.323$).

Similarly, for tumour free regions of the transition zone, the mean ADC was $1.336 \times 10^{-3}$ mm$^2$/s at time point 1, $1.243 \times 10^{-3}$ mm$^2$/s at time point 2, $1.551 \times 10^{-3}$ mm$^2$/s at time point 3 and $1.314 \times 10^{-3}$ mm$^2$/s at time point 4. No statistically significant differences were found between time points ($p=0.069$).

The distribution of ADC for lesion, normal PZ and TZ are shown in Figure 9-4.
**Figure 7–4** Trend of apparent diffusion coefficient (ADC) for 5 patients.

Time points are defined as: (1) – Before androgen deprivation therapy (ADT) (2)– After ADT, 3 weeks before external beam radiotherapy (EBRT) (3)– 6 weeks during EBRT (4)– 6 months after EBRT
7.3.4 FRACTIONAL INTRACELLULAR VOLUME

Before treatment, the mean lesion FIC was 0.621 (range 0.729-0.553) and median FIC was 0.607. After ADT, the mean lesion FIC decreased by 14% to 0.534 (p>0.99). At time point 3 (after 6 weeks of radiotherapy), the mean FIC decreased by 38% to 0.386 (p=0.086). After 6 months, FIC decreased further by 26% to 0.245 (p>0.99). There were statistically significant differences between the FIC before any therapy to FIC, six months after EBRT (p=0.009).

For tumour-free regions of the peripheral zone, the mean FIC before therapy was 0.074. The mean FIC after hormones was 0.087, 0.048 after 6 weeks of EBRT and 0.063 after 6 months of EBRT. This was no statistically significant difference between these time points (p=0.346). Similarly, for tumour free regions of the transition zone, the mean FIC was higher at 0.177 at time point 1, 0.203 at time point 2, 0.061 at time point 3 and 0.078 at time point 4. There was a statistically significant difference between mean FIC after hormones and FIC after 6 weeks of radiotherapy (p=0.029). No other significant differences were seen (p=0.668, p=0.086 and p>0.99).

The distribution of FIC for lesion, normal PZ and TZ are shown in Figure 9-5.
Figure 7-5 Trend of fractional intracellular volume (FIC) for 5 patients.

Time points are defined as: (1) – Before androgen deprivation therapy (ADT) (2)– After ADT, 3 weeks before external beam radiotherapy (EBRT) (3)– 6 weeks during EBRT (4)– 6 months after EBRT
Before treatment, the mean lesion FEES was 0.240 (range 0.089-0.302) and median FEES was 0.284. After ADT, the mean lesion FEES increased by 14% to 0.272 (p>0.99). At time point 3 (after 6 weeks of radiotherapy), the mean FEES increased by 52% to 0.413 (p=0.518). After 6 months, FEES increased further by 26% to 0.523 (p>0.99). There were statistically significant differences between the FEES before any therapy compared to FEES six months after EBRT (p=0.042) and between FEES after ADT and 6 months after EBRT (p=0.042).

For tumour-free regions of the peripheral zone, the mean FEES before therapy was 0.523. The mean FIC after hormones was 0.441, 0.495 after 6 weeks of EBRT and 0.586 after 6 months of EBRT. This was no statistically significant difference between these time points (p=0.145). Similarly, for tumour free regions of the transition zone, the mean FEES was 0.595 at time point 1, 0.443 at time point 2, 0.502 at time point 3 and 0.603 at time point 4. There were no statistically significant differences between FEES at the four different time points (p=0.095).

The distribution of FEES for lesion, normal PZ and TZ are shown in Figure 9-6.
Figure 7-6 Trend of fractional extravascular extracellular volume (FEES) for 5 patients.

Time points are defined as: (1) – Before androgen deprivation therapy (ADT) (2) – After ADT, 3 weeks before external beam radiotherapy (EBRT) (3) – 6 weeks during EBRT (4) – 6 months after EBRT
7.3.6 Fractional Vascular Volume

Before treatment, the mean lesion FVASC was 0.062 (range 0.04–0.073) and the median FVASC was 0.064. After ADT, the mean lesion FVASC increased by 31% to 0.080. At time point 3 (after 6 weeks of radiotherapy), the mean FVASC increased slightly to 0.096 (19%). After 6 months, FVASC was stable at 0.095 (1% decrease). There was no statistically significant difference between FVASC at any time point (p=0.188).

For tumour-free regions of the peripheral zone, the mean FVASC before therapy was 0.211. The mean FVASC after hormones was similar at 0.212 and increased slightly 6 weeks after radiotherapy to 0.240 and decreased to 0.179, 6 months after EBRT. This was no statistically significant difference between these time points (p=0.204). For tumour free regions of the transition zone, the mean FVASC was 0.201 at time point 1, 0.245 at time point 2, 0.278 at time point 3 and 0.302 at time point 4. There was no statistically significant difference in FVASC at the four different time points (p=0.724).

The distribution of FVASC for lesion, normal PZ and TZ are shown in Figure 9-7.
**Figure 7-7** Trend of fractional vascular fraction (FVASC) for 5 patients.

Time points are defined as: (1) – Before androgen deprivation therapy (ADT) (2) - After ADT, 3 weeks before external beam radiotherapy (EBRT) (3) - 6 weeks during EBRT (4) - 6 months after EBRT
Gland volume changed with ADT and radiotherapy. The mean gland volume before therapy was 28 mls (range 23-38 mls). After ADT, mean gland volume decreased by 14% to 24 mls. After 6 weeks of EBRT, gland volume decreased by 6% to 22.8 and a further 4% 6 months after EBRT to 22 mls. This change is demonstrated in Figure 9-8.

**Figure 7-8** Change in gland volume at the four defined time points.

Time points are defined as (1) – Before androgen deprivation therapy (ADT) (2)– After ADT, 3 weeks before external beam radiotherapy (EBRT) (3)– 6 weeks during EBRT (4)– 6 months after EBRT
7.4 DISCUSSION

In this study, men with biopsy-proven prostate cancer who opted for androgen deprivation therapy and radiotherapy underwent VERDICT MRI and multiparametric MRI at four time points. Serial changes in parameters derived from mpMRI and VERDICT MRI were compared.

The most dramatic change was seen with PSA and PSAD which decreased by 97% after hormonal therapy. This is an expected finding as PSA expression is directly related to activation of the androgen receptor (AR) by circulating androgens in normal prostate epithelium. Therefore, blocking AR activation leads to PSA suppression. The use of ADT also leads to decreased expression of AR-stimulated PSA in tumour cells, further lowering PSA. EBRT also lowered PSA but not by the same degree. Previous studies have shown that biopsies can be positive before PSA rise \[^{108}\] and the dramatic reduction in PSA after ADT, may make detecting small rises difficult.

ADC showed inconsistent changes in the five patients in this study, increasing in some after ADT and decreasing in others. After six months of radiotherapy, there was generally a small increase in tumour ADC. Due to the inconsistent changes after ADT and EBRT, there were no statistically significant changes. In tumour-free tissue, ADC similarly showed no consistent trend for ADT or radiotherapy.

In contrast, two VERDICT parameters (FIC and FEES) showed statistically significant changes between values before therapy and six months after radiotherapy. These changes could reflect the generalised atrophy seen in prostates after radiotherapy which takes time to develop \[^{109}\]. This finding has important implications as PSA shows dramatic changes to ADT but not EBRT. Therefore, a biomarker that shows a response to EBRT could be useful in monitoring treatment. This study also provides some guidance in terms of the timing of imaging when
monitoring EBRT treatment. Although changes are seen 6 weeks during EBRT, statistical differences were seen after 6 months post-EBRT. In tumour-free tissue, there were no statistically significant differences, although a generalised reduction in FIC and an increase in FEES was seen.

There were no significant changes in the vascular fraction as estimated by VERDICT MRI. This is consistent with the results discussed in previous chapters. One reason for this could be that changes in cellular architecture may dominate over changes in the vascular fraction. Another possible reason is that the VERDICT model may not accurately estimate the vascular fraction due to limitations of the model, for instance, fixed intrinsic diffusivity. It has been shown that the repeatability of FVASC is poorer compared to FEES and FIC [59].

One of the main limitations of this study is that only 5 men were analysed. Therefore, the generalisability of findings to the population studied may be limited. All men included in the study were responders to therapy and therefore we cannot assess the ability of VERDICT MRI in detecting recurrent or residual disease. Both VERDICT and mpMRI were performed on one scanner and therefore reproducibility on other scanners cannot be assessed.

Conclusion

In this pilot study, VERDICT MRI parameters showed a more consistent response to therapy compared to ADC. VERDICT MRI performed six months after EBRT may have some value in monitoring disease.
8 THESIS CONCLUSIONS AND FUTURE OUTLOOK

In this thesis, I evaluated whether quantitative imaging biomarkers derived from VERDICT MRI could help in characterising prostate cancer. In this chapter, I synthesise my findings and discuss the work that needs to be done to allow translation of biomarkers studied in this work.

My initial aim was to work out the current state of play in prostate cancer diagnosis and whether there is a clinical need for additional biomarkers. This aim was addressed by reviewing published literature on prostate cancer diagnosis, understanding the pathological basis of prostate cancer, and analysing the clinical outcomes from the INNOVATE study. The results showed that the overall application of mpMRI in the diagnostic pathway before biopsy can help spare men with normal MRI appearances (Likert 2 or below) an invasive biopsy. Some men who have an indeterminate appearance (Likert 3) can also be spared biopsy by applying PSA density thresholds. However, a large proportion of men (87%) who have Likert 3 who go on to have a biopsy, have a negative biopsy or insignificant prostate cancer. Furthermore, about half (49%) of men with Likert 4 lesions also have unnecessary negative biopsies. When radiologists score a lesion Likert 5, the likelihood of there being significant cancer approaches 100%.

These results likely reflect the heterogeneity of MRI appearances of tumours and the fact that due to some lesions not being conspicuous, radiologists tend to give indeterminate scores in order not to miss any cancer. In other words, specificity is sacrificed for sensitivity which makes sense clinically. However, in the cohort studied, three quarters (76%) of men had either a Likert 3 or Likert 4 score which results in a significant number of invasive biopsies. This analysis shows that there is a clinical need for biomarkers to better risk stratify these men.
After identifying this clinical need, I investigated which biomarkers from VERDICT MRI could add value. I compared these biomarkers to ADC and PSA density, which have been shown to have value in the literature. I found that fractional intracellular volume (FIC) and fractional extracellular and extravascular volume (FEES) values showed a greater differentiation in men who had csPCa to those that didn’t, compared to ADC and PSAD. This was also seen in subgroup analysis based on which scanner a participant was scanned on. The vascular fraction as estimated by VERDICT was not a good differentiator of disease and seemed to decrease with Gleason grade which was unexpected. To use these values clinically, I derived thresholds from the data which could help aid biopsy decisions in men who have Likert 3 or Likert 4 scores. As the thresholds were derived from the same data that they were then re-applied to, their performance may not be reflective of when applied to a different cohort. The next step would be to first use these thresholds in a different prospective cohort.

Image quality has a big impact on the detection of prostate cancer in men suspected of prostate cancer. I investigated what the image quality of FIC maps was when assessed by clinical radiologists who have not assessed these maps before. I compared the image quality of FIC to ADC which is used in clinical practice. Image quality was comparable for FIC maps when compared to ADC maps for both readers and showed similar trends for subgroups scanned on 1.5T or a 3T scanner. Inter-observer agreement for FIC and ADC maps was overall comparable when read individually and both increased when the overall dataset was read. Considering the unfamiliarity of readers to FIC maps, it is encouraging that observer agreement for FIC was similar to ADC maps. For FIC datasets, the agreement increased to substantial and was slightly lower for ADC datasets. With increased exposure to FIC maps, one can expect agreement rates to rise. The time taken to read FIC maps was not significantly longer compared to ADC for both readers in the study. This is a useful finding from a health economics point of view. Although a formal cost-effectiveness analysis is needed to compare VERDICT to mpMRI.
I also investigated whether the qualitative assessment of FIC maps could have an impact on the risk stratification of men in a multi-reader study. Although the sample size taken from the INNOVATE cohort was small (19%), there was a statistical difference in the number of false positives when readers assessed FIC maps compared to ADC maps individually. This effect was then diluted when readers scored the entire dataset which was likely due to indeterminate appearances on other sequences such as DCE.

Although the qualitative analysis is not complete, I propose that quantitative analysis is the better method for interpreting FIC maps. The advantages of QIBs are manyfold such as better reproducibility, reduced inter-observer variability, and the prospect of automated analysis and machine learning. For men on active surveillance with MRI, a QIB would a useful tool in tracking MRI lesions and deciding when to intervene. Currently, ADC has been used to monitor lesions, but thresholds have not been established in the literature. Qualitative assessment has been shown to be superior to ADC measurements \[99\]. This represents an opportunity for VERDICT MRI and work is underway within our group in testing VERDICT in the context of active surveillance. However before QIBs are translated clinically, they need rigorous repeatability and reproducibility studies which can be time-consuming. Reproducibility studies are also planned in collaboration with other sites.

The biomarker response to therapy is an important step in biological validation. I showed in a small cohort who had serial imaging that changes in VERDICT parameters reflected a patient’s response to hormones and radiotherapy. There remains a need for biomarkers to identify men who have recurrent or residual disease after radiotherapy. I showed in a meta-analysis (Appendix) that oncological outcomes can be poor for these men. A larger study is needed to see whether the residual disease can be characterised on VERDICT as in the cohort studied in this thesis, there were no such patients.
In conclusion, this body of work represents a step forward in the clinical translation of biomarkers derived from VERDICT by evaluating its clinical utility (Figure 10-1). Work is underway to measure reproducibility at other sites. Once this has been completed, I believe based on the findings of this thesis, there is enough justification to run a multi-centre, blinded, prospective trial using VERDICT MRI to make biopsy decisions, which would enable the jump over the translational gap.

**Figure 8-1** Imaging Biomarker roadmap for VERDICT MRI after this Thesis.
REFERENCES


[54] Panagiotaki E, Walker-Samuel S, Siow B, Johnson SP, Rajkumar V, Pedley RB, et


https://doi.org/10.1016/j.juro.2011.03.147.


Rosenkrantz AB, Oei M, Babb JS, Niver BE, Taouli B. Diffusion-weighted imaging of the abdomen at 3.0 Tesla: Image quality and apparent diffusion coefficient


https://doi.org/10.1097/JU.0000000000000617.


https://doi.org/10.1097/JU.0000000000000757.


https://doi.org/10.1016/j.eururo.2018.11.023.

https://doi.org/10.1259/bjr.20170645.


9 Appendix

9.1 SELECTED PUBLICATIONS ON PROSTATE IMAGING

9.1.1 CLINICAL OUTCOMES OF THE INNOVATE TRIAL


*Equal Contributors

Evaluation of PSA and PSA Density in a multiparametric magnetic resonance imaging-directed diagnostic pathway for suspected prostate cancer: the INNOVATE Trial

Hayley Pye, a,*, Saurabh Singh, b,c,*, Joseph M. Norris, a,d, Lina M. Carmona Echeverria, a,d, Vasilis Stavrinides, a,d, Alistair Grey, d,e, Eoin Dinneen, d,e, Elly Pilavachi, d,e, Joey Clemente, b,c, Susan Heavey, a, Urszula Stopka-Farooqui, a, Benjamin S. Simpson, a, Elisenda Bonet-Carne, f, Dominic Patel, g, Peter Barker, h, Keith Burling, h, Nicola Stevens, b,c, Tony Ng, i, Eleftheria Panagiotaki, f, David Hawkes, f, Daniel C. Alexander, f, Giorgia Trevisan, g, Manuel Rodriguez-Justo, g, Aiman Haider, g, Alex Freeman, g, Alex Kirkham, c, David Atkinson, b, Clare Allen, c, Greg Shaw, d,e, Teresita Beeston, b, Mrishta Brizmohun Appayya, b, Arash Latifoltojar, b,c, Edward W. Johnston, b,c, Mark Emberton, a,d, Caroline M. Moore, a,d, Hashim U. Ahmed, j,k, Shonit Punwani, b,c,#, and Hayley C. Whitaker,a,#,

a Molecular Diagnostics and Therapeutics Group, UCL Division of Surgery & Interventional Science, University College London, London, UK
ABSTRACT: Objectives: To assess the clinical outcomes of mpMRI before biopsy and evaluate the space remaining for novel biomarkers. Methods: The INNOVATE study was set up to evaluate the validity of novel fluidic biomarkers in men with suspected prostate cancer who undergo pre-biopsy mpMRI. We report the characteristics of this clinical cohort, the distribution of clinical serum biomarkers, PSA and PSA density (PSAD), and compare the mpMRI Likert scoring system to the Prostate Imaging–Reporting and Data System v2.1 (PI-RADS) in men undergoing biopsy. Results: 340 men underwent mpMRI to evaluate suspected
prostate cancer. 193/340 (57%) men had subsequent MRI-targeted prostate biopsy. Clinically significant prostate cancer (csigPCa), i.e overall Gleason≥3+4 of any length OR maximum cancer core length (MCCL) 4mm of any grade including any 3+3, was found in 96/195 (49%) of biopsied patients. Median PSA (and PSAD) was 4.7 (0.20), 8.0 (0.17), and 9.7 (0.31) ng/ml (ng/ml/ml) in mpMRI scored Likert 3,4,5 respectively for men with csigPCa on biopsy. The space for novel biomarkers was shown to be within the group of men with mpMRI scored Likert3 (178/340) and 4 (70/350), in whom an additional of 40% (70/178) men with mpMRI-scored Likert3, and 37% (26/70) Likert4 could have been spared biopsy. PSAD is already considered clinically in this cohort to risk stratify patients for biopsy, despite this 67% (55/82) of men with mpMRI-scored Likert3, and 55% (36/65) Likert4, who underwent prostate biopsy had a PSAD below a clinical threshold of 0.15 (or 0.12 for men aged <50 years). Different thresholds of PSA and PSAD were assessed in mpMRI-scored Likert4 to predict csigPCa on biopsy, to achieve false negative levels of ≤ 5% the proportion of patients whom who test above the threshold were unsuitably high at 86 and 92 % of patients for PSAD and PSA respectively. When PSA was re tested in a subcohort of men repeated PSAD showed its poor reproducibility with 43% (41/95) of patients being reclassified. After PI-RADS rescoring of the biopsied lesions, 66% (54/82) of the Likert3 lesions received a different PI-RADS score. Conclusion: The addition of simple biochemical and radiological markers (Likert and PSAD) facilitate the streamlining of the mpMRI-diagnostic pathway for suspected prostate cancer but there remains scope for improvement, in the introduction of novel biomarkers for risk assessment in Likert3 and 4 patients, future application of novel biomarkers tested in a Likert cohort would also require re-optimisation around Likert3/ PI-RADS2, as well as reproducibility testing.

Keywords: Biomarkers; INNOVATE; multiparametric MRI; Prostate Cancer; Diagnosis; PSA Density

INTRODUCTION

Prostate cancer represents a significant global healthcare challenge[3,110,111]. The traditional diagnostic approach of combining serum prostate specific antigen (PSA) testing with subsequent systematic transrectal ultrasound (TRUS)-guided
biopsy is now widely accepted as being suboptimal for identifying men at highest risk \[^{22,112}\]. The introduction of pre-biopsy multiparametric magnetic resonance imaging (mpMRI) has greatly enhanced the risk stratification of men with suspected prostate cancer, enabling the omission of biopsy in low-risk men, and use of image-guided biopsy in high-risk patients.\[^{22}\] This strategy reduces the number of prostate biopsies required, and results in a decrease in the overdiagnosis of clinically insignificant disease and an increase in the diagnosis of clinically significant disease.\[^{22,35,113,114}\] Almost all clinical guidelines now support pre-biopsy mpMRI in men with suspect prostate cancer; including; the National Institute for Health and Care Excellence (NICE), the European Association of Urology (EAU) and the American Urological Association (AUA) \[^{39,115–117}\]. A common biomarker used alongside mpMRI is MRI-derived PSA density (PSAD), particularly in the case of intermediate or negative mpMRI results, where a threshold of 0.15 ng/ml/ml is commonly advocated \[^{34,69,118}\]. The 2019 NICE guidelines suggest a PSA density threshold of 0.15 ng/ml/ml can be used as part of the decision to biopsy a men with raised PSA and negative MRI (Likert score of 1 or 2) \[^{39}\]. The 2019 EAU guidelines also discuss the utility of PSAD to identify patients that can safely avoid biopsy in case of a negative mpMRI, but include no formal cutoff in their final recommendations \[^{115}\].

The INNOVATE trial (ClinicalTrials.gov: NCT02689271; combIning advaNces in imagiNg with biOmarkers for improVed diagnosis of Aggressive prosTate cancer) is a prospective cohort study evaluating the value of additional fluidic biomarkers and a novel diffusion-weighted MRI technique to the mpMRI-directed diagnostic pathway for suspected prostate cancer.\[^{119}\] Here we present the clinical diagnostic outcomes of the mpMRI-directed pathway of the INNOVATE study and examine the distribution of mpMRI scores and PSA density (PSAD) in this cohort in relation to clinically significant cancer on biopsy. We evaluate what space remains for other serum and urine biomarkers, and compare the Likert and PIRADS version 2.1 scoring system in men with biopsy to better relate our findings to centres utilising this scoring method.
MATERIALS AND METHODS

Patient Population and Study Eligibility

The INNOVATE study protocol has been described in-depth elsewhere.[119] In brief, men referred with suspected prostate cancer between April 2016 to September 2019 underwent serum and urine collection, and pre-biopsy mpMRI (with additional VERDICT MRI (Vascular, Extracellular, and Restricted Diffusion for Cytometry in Tumours[55])), followed by MRI-targeted biopsy when indicated as part of usual clinical care. Inclusion and exclusion criteria for the full INNOVATE study are shown in (Figure S1) and published previously [119]. Initial radiological end points from the trial have also been published [59]. In line with the inclusion criteria, the pilot stage of the study included both men with and without a prior prostate cancer diagnosis, later recruitment stages omitted men with a prior positive biopsy in favour of men undergoing diagnosis. For this publication to better reflect a diagnostic cohort, the men with prior positive prostate biopsy (n=42) were excluded, alongside men for whom biopsy was recommended clinically but not performed (n=9), reasons for this included patient refusal / loss to clinical follow up (n=4), patient choice to have the biopsy elsewhere / in a private clinic (n=3) or biopsy was prevented for other clinical reasons (n=2) (Figure S1).

Multiparametric MRI

All men underwent a clinical prostate mpMRI at UCLH on either a 1.5T or 3.0 Tesla scanner (Achieva; Philips, Best, the Netherlands / Ingenia, Phillips, Best, the Netherlands / Avanto, Siemens, Erlangen, Germany) using a pelvic-phased array coil. mpMRI sequences included T1-weighted (T1W), T2-weighted (T2W), dynamic contrast enhancement (DCE) with gadolinium, diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) maps. The additional VERDICT sequences were for research only and so not described herein or used as part of the clinical decision making for these patients. All mpMRIs were scored at UCLH as part of standard clinical practice by experienced uroradiologists including (AK, CA, SP) using a five-point ordinal Likert scale for the likelihood of clinically significant prostate cancer[69]. The ‘Likert score’ is a cognitive score based on radiologist experience after inspection of the prostate using different sequences and interpreted alongside clinical data available to them at referral. If multiple
suspicious areas are present, these are scored separately. The Likert score reported in this analysis represents the highest score in any area of the prostate given to a patient. For this analysis, men who underwent biopsy also had their highest scoring lesion on mpMRI re-scored using the Prostate Imaging–Reporting and Data System v2.1 (PI-RADS) scoring system [116], this secondary reporting was done retrospectively and not used to influence clinical decision making. Blinding of lesion Likert score for re-reporting was not done due to the rigid nature of PI-RADS statements and because it was required to identify the right lesion for rescoring. Men who did not undergo biopsy were not re-scored with PI-RADS. This was because of ethical implications as well as a lack of histopathological outcome which would allow a meaningful comparison between the two scoring systems.

Biopsy

INNOVATE is a non-interventional trial and so decision to biopsy in this cohort reflects clinical practice and the current NICE guidelines, i.e. each patient received an individual decision to biopsy based on mpMRI results, PSA / PSAD level as well as other clinical information if it is available including (but not limited to); PSA velocity, DRE findings, family history, other risk factors and comorbidities as well as life expectancy. The decision is made in discussion with the patient and so patient choice is also a key factor. In this cohort prostate biopsy was performed a median of 28 days after mpMRI (IQR: 18-50). The majority (81%, 158/195) of biopsies were carried out at UCLH, these biopsies were taken transperineally and consisted of MRI-targeted deployments only (cognitive method) with 1% (2/158) of men having additional systematic sampling if indicated clinically. Reflecting real-world variation in clinical practice 19% (37/195) of men in this cohort had their care transferred to a sister hospital (Barts Health), biopsies carried out here were taken transrectally and the majority (76%, 28/37) consisted of MRI-targeted deployments (cognitive method) with additional systematic sampling (Table S1). Clinically significant prostate cancer (csigPCa) is defined as overall Gleason ≥3+4 of any length OR maximum cancer core length (MCCL) ≥4mm of any grade, concordant with definition 2 used in the PROMIS validation report [22]. These criteria were previously developed and validated for detection of Gleason score 4 or greater and cancer core lengths representative of lesions 0.5 mL or greater[68]. Data broken down by other definitions are available as supplementary figures.
Clinical Serum PSA and PSA Density (PSAD)

INNOVATE recruited at secondary care level from a population of men who were referred with a ‘suspicion for prostate cancer,’ so accordingly 82% (283/340) of the men in this cohort had a referral PSA level above the age adjusted referral limit (> 3.0 ng/mL for men aged 50-69 years old, and > 5.0 ng/mL for men at or over 70 years old [70]). The clinical serum PSA measurements were taken from the clinical notes and PSA testing was carried out in primary or secondary care prior to referral or around the time of mpMRI. The median time between PSA measurements and mpMRI in this cohort was -22 days (IQR -49 to 0 days). The median time between PSA measurement and prostate biopsy (for those who had one) was 33 days (IQR 9 to 68 days). PSA density (PSAD) was calculated by dividing serum PSA (ng/mL) by mpMRI-derived prostate volume (mL) measured using the prolate ellipsoid method (width x length x height x 0.52) [69].

‘Lab’ PSA measurement

For the figure describing reproducibility of PSAD new serum samples were collected at the point of patient mpMRI for analysis of total serum PSA. Samples were taken a median of 13 days after the patient’s clinical PSA test was done and all were within 3 months (IQR = -minus 39 to plus 13 days). All samples were sent for re-testing at an external clinically accredited facility; CBAL (Core Biochemical Assay Laboratory) at Cambridge University Hospitals NHS Foundation Trust. The following assay was used; Total PSA (Beckman Coulter Access Hybritech PSA (Ref 37200) using a Beckman Coulter Access 2 immunoassays analyser. Total PSA in ng/ml was measured directly in patient serum and made comparable to a standard curve. The acceptance of each batch was undertaken by running quality control samples at the beginning and end of each assay, with the analysis being accepted only when the quality controls were within the predefined limits. The clinical and the repeated lab PSA were positively correlated with an rho² of 0.4 (p = 3x10^-10). There was no correlation between the difference between the two values and time between the two tests (r² of 0.006 (p = 0.5) (Figure S2).

Statistical Analyses

All statistical analyses were performed in RStudio version 1.3.1056 within the R statistical environment v4.0.2. Comparison of median tests used include;
Independent 2-group Mann-Whitney U Test (stats package v4.0.2 wilcox.test) and Kruskal-Wallis Rank Sum Test with Dunn's Kruskal-Wallis Multiple Comparisons (stats package v4.0.2 kruskal.test and FSA package 0.8.30 dunnTest) these tests use PSA or PSAD as numeric input data and grouping via ordered factors (either MRI score or biopsy outcome). For correlation tests PSA or PSAD values were log2 transformed before correlation using Spearman's rank correlation (two.sided, exact = FALSE) (stats package v4.0.2 cor.test).
RESULTS

Clinical outcomes

In this cohort, 340 men underwent pre-biopsy mpMRI for prostate cancer diagnosis, of these 57% (195/340) had a subsequent MRI-targeted prostate biopsy (Table 1). The proportion of men who had biopsy increased with mpMRI Likert score; 4% (2/47) of men with Likert score 2, 46% (82/177) of men with Likert score 3, 93% (65/70) of men with Likert score 4, and 100% (46/46) of men with Likert score 5 had a biopsy (Table 1).

Table 1

<table>
<thead>
<tr>
<th>Patients did NOT have a biopsy (n=45)</th>
<th>Patients had a biopsy (n=195)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-biopsy mpMRI:</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>Referral PSA (ng/ml)</td>
<td></td>
</tr>
<tr>
<td>Likert score 2</td>
<td></td>
</tr>
<tr>
<td>Likert score 3</td>
<td></td>
</tr>
<tr>
<td>Likert score 4</td>
<td></td>
</tr>
<tr>
<td>Likert score 5</td>
<td></td>
</tr>
<tr>
<td>PSA Density (ng/ml)</td>
<td></td>
</tr>
<tr>
<td>mg/ml</td>
<td></td>
</tr>
<tr>
<td>Likert score 2</td>
<td></td>
</tr>
<tr>
<td>Likert score 3</td>
<td></td>
</tr>
<tr>
<td>Likert score 4</td>
<td></td>
</tr>
<tr>
<td>Likert score 5</td>
<td></td>
</tr>
<tr>
<td>MRI volume (ml)</td>
<td></td>
</tr>
<tr>
<td>Likert score 2</td>
<td></td>
</tr>
<tr>
<td>Likert score 3</td>
<td></td>
</tr>
<tr>
<td>Likert score 4</td>
<td></td>
</tr>
<tr>
<td>Likert score 5</td>
<td></td>
</tr>
<tr>
<td>Biopsy Result</td>
<td></td>
</tr>
<tr>
<td>Negative biopsy</td>
<td></td>
</tr>
<tr>
<td>Positive biopsy</td>
<td></td>
</tr>
<tr>
<td>Biopsy MCCL (mm)</td>
<td></td>
</tr>
<tr>
<td>Statistics presented: median (IQR); n (% of column)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. – Summary demographic, clinical, radiological and pathological outcome data for all men included in the INNOVATE trial, stratified by highest Likert score. Grouped by (Table A) decision to biopsy and (Table B) presence of clinically significant prostate cancer on biopsy. Biopsy result reported as prognostic grade groups (PGG). Clinically significant prostate cancer (csigPCa) is defined as overall

We have defined clinically significant prostate cancer (csigPCa) as overall Gleason ≥3+4 of any length or maximum cancer core length (MCCL) ≥4mm of any grade (see methods). Of the 195 men who had a prostate biopsy csigPCa of this definition was diagnosed in 49% (96/195), rising to 56% (110/195) for cancer of any grade or size (Table 1). As a proportion of men biopsied in each Likert group, the number of men who had csigPCa increased with Likert score; 0% (0/2) of men with Likert 2, 15% (12/82) of men with Likert score 3, 60% (39/65) of men with Likert score 4, and 98% (45/46) of men with Likert score 5 (Table 1). The number of men in these groups above and below a PSAD threshold of 0.15 (or 0.12 for men aged <50) is shown in (Figure 1) and discussed in more detail for Likert 3 and 4 men below.

Figure 1

![Figure 1](image-url)

**Figure 1.** – Men characterised as above or below a PSAD threshold of 0.15ng/ml/ml or 0.12 for men under 50 years old. Stratified by score on MRI. (A) Men who did not undergo prostate biopsy (grey). (B) Men who had a subsequent prostate biopsy grouped as follows; 1. No cancer was found (Blue), 2. Clinically significant prostate cancer was found (Yellow), 3. Non-significant cancer was found (Green). Clinically significant prostate cancer is defined as overall Gleason ≥3+4 of any length or maximum cancer core length (MCCL) ≥4mm of any grade. (C) Same
data replotted after highest scoring mMRI lesion was rescored retrospectively using PI-RADSv2.

For the men who underwent biopsy the mpMRI lesion that scored highest was re-scored with PI-RADS retrospectively. After PI-RADS rescoring, 60% of lesions had the same score as their Likert score (117/195), of the remaining lesions 11% (21/195) had a higher PI-RADS score and 29% 57/195 had a lower PI-RADS score, no lesions underwent more than one category change up or down the scale (Table 2). After rescoring the proportion of patients in each PI-RADS group that had csigPCa on biopsy is also shown alongside Likert in (Figure 1 and Table 2).

However, because Likert was used to select for biopsy in this cohort the absolute proportions for each PI-RADS group that would have been selected for biopsy, or have had csigPCa on biopsy cannot be known.

Table 2

<table>
<thead>
<tr>
<th>Likert Score</th>
<th>PSAD Threshold</th>
<th>Below</th>
<th>Above</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likert2</td>
<td>N = 47^t</td>
<td>37</td>
<td>8</td>
</tr>
<tr>
<td>Likert3</td>
<td>N = 177^t</td>
<td>84</td>
<td>11</td>
</tr>
<tr>
<td>Likert4</td>
<td>N = 70^t</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Likert5</td>
<td>N = 46^t</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

^t Statistics presented: n

* PSAD threshold of 0.15 ng/ml/mL (or 0.12 if patient is younger than 50 yrs)

Table 2. For the men who underwent biopsy the mpMRI lesion that scored highest was re-scored with PI-RADS retrospectively. Proportion of patients with equivalent or a different PI-RADS score stratified by Likert score and having a PSAD above or below a threshold of 0.15 (or 0.12 for men aged <50). PI-RADS was not used to select patients for biopsy. Clinically significant prostate cancer is defined as overall Gleason \( \geq 3+4 \) of any length or maximum cancer core length (MCCL) \( \geq 4 \)mm of any grade.

Overall biopsy outcome of Likert 2 and 5 patients is shown in (Table 1 and Figure 1), however due to the high proportion of patients in these groups without and with
csigPCa respectively, discussion on the space for novel serum and urine biomarkers will focus on Likert 3 and 4 men. The detail on the Likert 2 and 5 men can still be read in supplementary text 1.

Likert 3

The number of Likert 3 men who underwent subsequent biopsy in our cohort was 46% (82/177) (Table 1). PSAD was considered clinically in the decision to biopsy, despite this it is clear other clinical variables are playing a role in the decision making because 67% (55/82) of the Likert 3 men who underwent prostate biopsy had a PSAD below the threshold of 0.15 (or 0.12 for men aged < 50) (Table S2). For the men who were below this threshold, the proportion that had csigPCa on biopsy was 9% (5/55), compared to 26% (7/27) of men above the threshold (Figure 1, Table S2). This data for other definitions of csigPCa and other PSAD thresholds is available in supplementary (Figure S6).

After PI-RADS rescoring of the biopsied Likert 3 lesions 66% (54/82) of the lesions received a different level of PI-RADS score compared to Likert; with 12% (10/82) receiving a PI-RADS 4 and 54% (44/82) receiving a PI-RADS 2 (Table 2). This large number of reclassified lesions reflects that Likert 3 (as used in UCLH) is a clinical definition that encompasses more than just the prescribed mpMRI features prescribed in PI-RADS. If the Likert 3 men that received a PI-RADS 2 score had not been biopsied a further 13% (44/340) of men could have been spared biopsy, however 4.5% (2/44) of these men had csigPCa on biopsy, increasing to 14% (6/44) if cancer of any definition is used (Figure 1 and 2). The two men that had csigPCa on biopsy had PSAD (PSA) values of 0.14 (6.1) 0.14 and 0.19 (4.2), one above the clinical PSAD threshold and one below. Comparatively of all the Likert 3 men with PI-RADS 2 lesions (n=44), 30% (13/44) would have also had a PSAD above the threshold. The median (IQR) PSAD for these men was 0.13 (0.10-0.15), significantly higher than the PSAD median (IQR) for all Likert 2 men; 0.09 (0.06-0.12) (p = 0.003), suggesting either this influenced their characterisation as Likert 3, or some feature of mpMRI (yet to be included in the PIRADS definition) associates with a higher PSAD (Table S2, S3). The full distribution of PSA and PSAD values for all men with Likert 3 can be seen graphically in (Figure S3, S4 and S5).

Likert 4
In the Likert 4 group, 5 men did not have a biopsy (5/70), 4 out of these 5 men had very low PSAD (PSA) values as follows; 0.04 (14), 0.06 (7.54), 0.05 (4.5), 0.02 (0.68), and the remaining man was 0.12 (7.85). The median (IQR) PSAD of all Likert 4 men was 0.13 (0.10-0.23) so the majority of the men spared biopsy were outliers in the lowest quartile of PSAD values. The full distribution of PSA and PSAD values for all men with Likert 4 can be seen graphically in (Figure S3, S4 and S5). Of the remaining men who did have biopsy, 55% (36/65) of these men had PSAD values below the PSAD threshold of 0.15 (or 0.12 for men aged <50) (Table S2). In the men below the PSAD threshold the proportion that had csigPCa on biopsy was 53% (19/36), compared to 69% (20/29) of men above the threshold (Figure 1, Table S2). This data for other definitions of csigPCa and other PSAD threshold is available in supplementary (Figure S7).

Since almost all men with Likert 4 were biopsied (65/70) we carried out confusion matrix analysis to see how good a range of PSA and PSAD thresholds would have been in this cohort to determine if Likert 4 men could avoid biopsy (Figure 2). In this context it is important to have low numbers of false negatives, to ensure every man with csigPCa gets a biopsy. However, to achieve a false negative value of 5% the false positive rate was 31%, if this threshold (0.09 ng/ml/ml) was applied to out cohort, 86% of all men in the group would be above it (Figure 2). Interestingly a similar predictive ability could be obtained with a PSA value above 4; i.e. a false negative value of 6% and false positive rate was 34% suggesting better biomarkers would be useful in this context. It should be noted that PSAD outperforms PSA at lower thresholds, i.e. with a greater false negative rate (Figure 2). Other definitions of cancer are also available in supplementary (Figure S8 and S9).
Figure 2. – Utility of PSA and PSAD thresholds in patients scored as Likert 4. A range of thresholds of PSA (A) or PSAD (B) were used to predict the presence of clinically significant prostate cancer on biopsy (Negative = no csigPCa, Positive = yes csigPCa). Clinically significant prostate cancer is defined as overall Gleason ≥3+4 of any length or maximum cancer core length (MCCL) ≥4mm of any grade. Each bar represents how accurate each threshold would be if used as a clinical test to predict for csigPCa on biopsy: e.g. for the highest PSAD threshold 0.18 (top bar graph B): if a patients PSAD was < 0.18 they would be ‘negative’ on the test and they would have around a 1 in 2 chance of cancer (as True Negative and False Negative are 31% and 31%), if their PSAD was > 0.18 they would be ‘positive’ on the test and they would have around a 3 in 4 chance of cancer (as False Positive and True Positive are 9% and 29% respectively). In this context we potentially selecting men for biopsy so would need a very low false negative, e.g. if we use the lowest PSAD threshold of 0.09 we get a 5% false negative rate and the highest true positive rate (55%) but 86% of all men tested would get a positive test result (False Positive plus True Positive) limiting the utility of the test. This data is for Likert 4 men only.

When Likert 4 lesions were rescoring with PI-RADS, 4.6% (3/65) were rescored as PI-RADS 3 and 17% (11/65) were rescored as PI-RADS 5. Of the 3 downgraded to PI-RADS 3, 60% (2/3) had csigPCa on biopsy. Of the 11 upgraded to PI-RADS 5 90% (10/11) had csigPCa on biopsy (Figure 1, Table 2).

PSA and PSAD reproducibility

To investigate the effect of PSA reproducibility on this cohort; total serum PSA was re-tested in the first 95 patients, serum samples were collected at the point of mpMRI and sent for re-testing at an external clinically accredited facility. When the repeated PSA values were converted to PSA density (PSAD) using the prostate volume as measured on mpMRI; 57% (54/95) maintained their classification as above or below a PSAD of 0.15, but 43% (41/95) would have been reclassified (Figure S10). This demonstrates the importance of repeatability for any new biomarkers, or panels of biomarkers aiming for implementation in this space.

CONCLUSIONS
Here we present the key diagnostic outcomes of the INNOVATE trial, reflecting real-life clinical outcome measures from experienced secondary care teams using a mpMRI-directed clinical pathway for men with suspected prostate cancer. Early experience at UCLH of using pre-biopsy mpMRI and PSAD to guide the decision to perform biopsy, 49% of men in this cohort were spared biopsy. Although not directly comparable, this is in line with previously published studies that use mpMRI alone to estimate the proportion of men who can be spared biopsy; 27% of men (derived from the PROMIS study[22]) or 50% of men (3T in bore biopsy study [120]). In our cohort, 47% of men indicated for biopsy were diagnosed with PGG 2 or higher, this is similar to the MRI-targeted biopsy arm of the PRECISION trial (38%) [35]. Any differences in patient numbers between our and historic UCLH cohorts cannot be reliably interpreted because over time UCLH has built on years of experience which has caused changes in its decision making processes and/or biopsy technique, and because all patients in this analysis were consented to a clinical trial and the recruitment or consenting process itself can self-selected for a certain group of patients.

In summary there remains space for novel serum and urine biomarkers in this new pathway. Around 35% (51/147) of men with Likert 3 or 4 disease on mpMRI were still suspicious for cancer and underwent prostate biopsy with no resulting clinically significant cancer diagnosis. The PSAD level of 0.15ng/ml/ml used as a trigger for biopsy in men with negative MRI (Likert 1 and 2) did not perform well enough in this cohort to help risk stratify patients of other Likert scores. Post hoc testing of other thresholds in the Likert 4 subgroup shows some patients could be spared but to get sufficient sensitivity the threshold would have to be so low almost all men would still be indicated for biopsy. Novel biomarkers in this space would aim to spare men biopsy, so a low false negative rate would be important.

When biopsied lesions were rescored with PIRADS, Likert 3 lesions received the lowest number of unchanged scores, something seen previously in a different cohort from the same center by Brizmohun and colleagues [121]. This suggests any novel biomarkers developed and validated in a Likert cohort would require particular re-optimisation around PI-RADS 2 and 3 before application to a PI-RADS cohort. Some significant cancers would have been missed in the PI-RADS 2
lesions that underwent biopsy in our cohort, so novel biomarkers could also have utility here. Pagniez et al discuss the utility of PSAD for mpMRI invisible cancers like these in a recent meta-analysis [118]. In a Likert based cohort from UCLH (like described herein) Norris et al have previously shown all invisible disease (Likert 1 and 2) were clinically insignificant [122].

Our comment of the utility provided by PSA and PSAD for prediction of clinically significant disease in specifically Likert 3 patients was limited by the unblinding of clinicians to these markers, as well the fact 55% (95/177) of men with Likert 3 disease did not undergo biopsy, and so had no confirmed histopathological outcome. Because the clinical decision to biopsy (and Likert scores) were already influenced by PSA and PSAD, any analysis of biopsy outcome is potentially subject to selection bias. However this question has been comprehensively answered by others in a study with template biopsy for all patients as the reference standard. In a post-hoc analysis of the PICTURE trial, men with indeterminate (i.e. Likert 3) lesions and clinically significant disease on biopsy were found to have significantly higher PSAD compared to men without significant disease (0.19 vs. 0.13; p = 0.004), and a threshold of >0.17 ng/ml was shown to improve prediction of significant cancer on biopsy, but 9% of patients with significant cancer would be missed [121]. This aligns in our cohort; in the biopsied Likert 3 men there was a difference in the median PSAD between men with and without csigPCa; 0.12 (IQR 0.08-0.15) and 0.20 (IQR 0.13-0.26) (p = 0.02) the higher level being associated with csigPCa (Figure S3).

The repeatability of total PSA has been studied elsewhere [123]. In the Roehrborn study retrospective analysis of clinical data from one centre showed one third of the patients had a difference of greater than +/- 1.0 ng/mL on repeat PSA. In our study this value was higher at 85% of patients. The additional variability is likely due to the fact samples were measured in different clinical testing facilities and most likely using different testing platforms. If an absolute threshold for clinical action is to be implemented with PSAD, the reproducibility of PSA testing needs to be addressed more fully. This problem with reproducibility would have to be
something other biomarkers or larger panels of biomarkers should aim to improve on.

Other potential limitations of the INNOVATE cohort in general are discussed in Supplementary text 2. The INNOVATE trial gives a clear, real-world illustration of a successful mpMRI-directed diagnostic pathway, and any samples and data acquired during this trial will provide a useful resource to help test novel biomarker for risk stratification of men with suspected prostate cancer. Future biomarker analyses of the INNOVATE cohort will focus on validation of novel and established blood and urine fluidic biomarkers, used at the point of mpMRI and the assessment of the utility of these panels to stratify and refine the mpMRI-directed pathway for suspected prostate cancer.

Funding and acknowledgements: This work is supported by Prostate Cancer UK: Targeted Call 2014: Translational Research St.2, project reference PG14-018-TR2. The Molecular Diagnostics and Therapeutics Group is also supported by Prostate Cancer UK and the Prostate Cancer UK Centre of Excellence which also funds LMCE. SH is funded by a Prostate Cancer UK Travelling Fellowship and JMN and VS are supported by Medical Research Council fellowships. BSS receives funding from the Rosetrees Trust. SP receives sessional funding from UCLH BRC. HUA research is supported by core funding from the United Kingdom's National Institute of Health Research (NIHR) Imperial Biomedical Research Centre and he currently receives funding from the Wellcome Trust, Prostate Cancer UK, The Urology Foundation, BMA Foundation, and Imperial Healthcare Charity. CMM receives research funding from PCUK, Movember, the Medical Research Council and the European Association of Urology Research Board. ME receives research support from the United Kingdom’s National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre for which he was awarded NIHR Senior Investigator status in 2013. AK is funded by the BRC.

Ethics: The study abides by the principles of the Declaration of Helsinki and the UK Research Governance Framework version 2. INNOVATE received UK Research Ethics Committee approval on 23rd December 2015 by the NRES Committee London—Surrey Borders with REC reference 15/LO/0692.
Department of Health Disclaimer: The views and opinions expressed therein are those of the authors and do not necessarily reflect those of Prostate Cancer UK, the NHS or the Department of Health.

Conflicts of interest: DH is a founder Shareholder of IXICO Plc. and adviser and shareholder for VisionRT. HUA’s research currently receives support from, Sonacare Inc., Trod Medical and Sophiris Biocorp for trials in prostate cancer. HUA is a paid medical consultant for Sophiris Biocorp and Sonacare Inc. and a proctor for Boston Scientific for Rezum and cryotherapy. ME acts as a consultant/adviser to Sonacare Medical and Exact Imaging. HUA and CMM are proctors for HIFU with Sonacare Inc. and paid for training other surgeons in this procedure. CMM has received consultancy and speaker fees from Astellas and Janssen, and research funding from Spectracer.

Autor contributions: Hayley Pye: Patient consent, data collection, figure production, paper writing, review & editing, project overview, project input: Saurabh Singh: Data creation, collection, figure production, paper writing, review & editing: Joseph M. Norris: figure production, paper writing, review & editing, project input: Lina M. Carmona Echeverria: figure production, paper writing, review & editing, project input: Vasilis Stavrinides: figure production, paper writing, review & editing, project input: Alistair Grey: data collection,: Eoin Dinneen: data collection,: Elly Pilavachi: data collection: Joey Clemente: recruitment, patient consent,: Susan Heavey: Patient consent, project input, review & editing: Ula Stopka-Farooqui: Patient consent, project input: Benjamin S. Simpson: project input, review & editing: Elisenda Bonet-Carne: designed clinical trial protocol, review & editing: Dominic Patel: designed clinical trial protocol: Peter Barker: Experimental work: Keith Burling: Experimental work, designed clinical trial protocol, review & editing: Nicola Stevens: designed clinical trial protocol: Tony Ng: designed clinical trial protocol: Eleftheria Panagiotaki: designed clinical trial protocol, review & editing: David Hawkes: designed clinical trial protocol, review & editing: Giorgia Trevisan: designed clinical trial protocol: Manuel Rodriguez-Justo: designed clinical trial protocol, review & editing: Aiman Haider: project overview, project development: Alex Freeman: project overview, project development: Alex Kirkham: designed clinical trial protocol, review & editing: David Atkinson: designed clinical trial protocol, review & editing: Clare Allen:
designed clinical trial protocol: Greg Shaw: recruitment, patient consent, Teresita Beeston: recruitment, project overview, project development, review & editing; Mrishta Brizmohun: recruitment, project overview, project development, review & editing; Arash Latifoltojar: recruitment, patient consent, project overview, project development; Edward W. Johnston: Initial ethic application, recruitment, patient consent, project overview, project development, review & editing; Mark Emberton: project overview, project development, review & editing; Caroline M. Moore: project overview, project development, review & editing; Hashim U. Ahmed: project overview, project development, review & editing; Shonit Punwani: grant application, project overview, project development, review & editing; Hayley C. Whitaker: grant application, project overview, project development, review & editing;

SUPPLEMENTARY MATERIALS: Figure S1: Flow diagram for patient recruitment in the INNOVATE trial, and a table detailing the inclusion and exclusion criteria for the full trial. Table S1: Further information about the prostate biopsies performed in this cohort. Figure S2: Total serum PSA re-tested in 95 patients. Supplementary text 1: Clinical outcome for Likert 2 and 5. Figure S3: PSA density of all patients, grouped as 1. without prostate biopsy, 2. without csigPCa on subsequent biopsy and 3. with csigPCa on biopsy. Figure S4: PSA of all patients, grouped as 1. without prostate biopsy, 2. without csigPCa on subsequent biopsy and 3. with csigPCa on biopsy. Figure S5: PSA density values for men plotted separately across their Likert and PI-RADS scores. Table S2: Proportion of patients above and below some clinically used PSA and PSAD thresholds. Figure S6: Proportion of biopsied Likert 3 patients above and below some clinically used PSA and PSAD thresholds. Figure S7: Proportion of biopsied Likert 4 patients above and below some clinically used PSA and PSAD thresholds. Figure S8: Confusion matrix outputs of a range of thresholds of PSAD used to predict the presence of clinically significant prostate cancer (various definitions shown). Figure S9: Confusion matrix outputs of a range of thresholds of PSA used to predict the presence of clinically significant prostate cancer (various definitions shown). Figure S10: Total serum PSA re-tested in 95 patients and converted to PSA density using the prostate volume as measured on mpMRI. Supplementary text 2: Other potential limitations of the analysis.
REFERENCES


https://doi.org/10.1073/PNAS.94.7.3320.

https://doi.org/10.1126/science.123.3191.309.


https://uroweb.org/guideline/prostate-cancer/.


[25] Kim CK, Park BK, Kim B. High-b-value diffusion-weighted imaging at 3 T to detect prostate cancer: Comparisons between b values of 1,000 and 2,000


Crook J, Malone S, Perry G, Bahadur Y, Robertson S, Abdolell M.


[116] Bjurlin MA, Carroll PR, Eggener S, Fulgham PF, Margolis DJ, Pinto PA, et al. Update of the Standard Operating Procedure on the Use of


9.1.2 Meta-analysis of biopsy outcomes in patients treated with EBRT

Publication discussed in Chapter 8 and 9 of the thesis, with the following citation:


LONG-TERM BIOPSY OUTCOMES IN PROSTATE CANCER PATIENTS TREATED WITH EXTERNAL BEAM RADIOTHERAPY: A SYSTEMATIC REVIEW AND META-ANALYSIS

Running Title:

BIOPSY OUTCOMES FOR RADIONECURRENT PROSTATE CANCER

Saurabh Singh¹, Caroline M Moore²,³, Shonit Punwani¹, Anita V Mitra⁴, Steve Bandula ¹,⁵
1 Centre of Medical Imaging, University College London, London, UK

2 Division of Surgery and Interventional Science, University College London, UK

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

4 Cancer Services, University College London Hospitals NHS Foundation Trust, London, UK

5 Interventional Oncology Service, University College London Hospitals NHS Foundation Trust, London, UK
INTRODUCTION

External beam radiotherapy (EBRT) has been a long-standing, recommended primary treatment for prostate cancer (PCa). With advances in delivery and treatment planning, EBRT has become safer, more precise, and more effective [1]. Despite these improvements, traditionally defined low- and intermediate-risk PCa patients treated with modern EBRT can still expect biochemical recurrence at 10 years in 10% and 23% of cases, respectively [2]. In addition to identifying recurrence as early as possible, it is also important to establish its location. PCa that has metastasised has different management options compared with radiorecurrent disease that is still confined to the prostate [3,4].

Biochemical failure (BCF) after EBRT is the most established predictor of PCa recurrence. BCF is currently assessed using the Phoenix criteria, defined as a rise by 2 ng/ml or more above nadir prostate specific antigen (PSA) levels [5]. Despite its obvious benefits, the primary drawback of BCF is that it does not discriminate between locally recurrent disease and metastasis. Increasingly for the assessment of local tumour control, imaging and biopsy are used [4]. Both multiparametric magnetic resonance imaging (mpMRI) and prostate specific membrane antigen-positive emission tomography (PSMA-PET) have shown promise in detecting residual or recurrent cancer after EBRT but can be difficult to interpret due to treatment related changes in the prostate and spatial resolution [6-8]. Therefore, to confirm recurrence or residual disease, biopsy is usually performed.

Biopsy has also been investigated as a stand-alone technique to evaluate local PCa recurrence after EBRT, albeit much less frequently [10]. Post-EBRT biopsies taken less than two years after EBRT treatment are not reliable and can still be challenging to interpret, for instance when there is cancer and treatment effect leading to ‘indeterminate’ reports alongside distinctly positive and negative [11]. Nevertheless, positive biopsy after EBRT implies failure of local tumour control and is associated with a downward trend in PCa prognosis [12], and in certain cases a positive biopsy can be found before BCF appears [13]. While a post-EBRT positive biopsy cannot rule out metastasis, it does provide direct histological evidence of local disease which could be targeted with salvage therapies.

219
A better understanding of the positive biopsy rate after EBRT also offers an opportunity to compare EBRT to other treatment modalities. To properly assess oncological efficacy, it is often expected that any new curative PCa treatment performs a post-treatment biopsy to verify the absence of cancer. Understanding the positive biopsy rate after EBRT will provide a valuable measuring stick.

In this meta-analysis we systematically reviewed the relevant literature and determined the post-EBRT positive biopsy rate when the biopsy was performed at least two years after treatment. We also assessed the associated risk of a positive biopsy vs. a negative biopsy as an indicator for long-term PCa outcome.

METHODS

Literature search and study selection

We searched for studies that utilized EBRT alone or in combination with androgen deprivation therapy (ADT) as primary treatment for low to intermediate risk PCa (PSA ≤ 20ng/ml, Gleason score ≤ 7, clinical stage ≤ T2b), where a biopsy at ≥ 2 years post-EBRT was an endpoint or observation of the study. Studies with high-risk PCa were accepted if they also included low- or intermediate-risk PCa in their patient population. Brachytherapy which is often combined with EBRT and proton beam therapy were not included, in order to focus on EBRT.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were utilized to search PUBMED/MEDLINE and EMBASE. We defined study eligibility with reference to Population, Intervention, Outcome, and Study design. A structured literature search for studies until March 2020 with keywords "prostatic neoplasms"; "prostate cancer"; "biopsy"; "radiotherapy"; and not "brachytherapy" was conducted. The separate database searches were imported into Mendeley Desktop (Mendeley, London, United Kingdom) to detect and remove duplicates. Search results containing editorials, review, case-reports, models and opinion-pieces were automatically removed from consideration. Thereafter, the remaining search results were screened by two authors (SS and SB) for relevant keywords in the title and abstract. Only randomised controlled trials or
cohort studies published in English journals were used. In cases where two or more studies reported results of an overlapping patient cohort, the one with higher number of patients biopsied at ≥ 2 years was selected. In one case the overlapping patient cohort (Zelefsky et al 2018 and Zelefsky et al 2019) was clarified by contacting the study author directly.

Data extraction

A variety of information was extracted by the same two authors (SS and SB) from each eligible study, including: PCa risk group breakdown, EBRT technique, dose, number of patients who received a biopsy ≥ 2 years after EBRT, and the positive/indeterminate biopsy rate. The baseline PCa risk stratification was not consistent across studies, including the clinical T stage, Gleason score and National Comprehensive Cancer Network (NCCN) risk assessments. If available, the 5 to 10-year long-term data on PCa progression such as biochemical failure (BCF), distant metastases-free survival (DMFS), and prostate cancer-specific mortality (PCSM) was also recorded. If high-risk PCa was also included in the study population, the number of patients counted, the post-EBRT positive biopsy rate and any long-term outcomes were recorded for only the low- to intermediate-risk group when possible, although some studies did not differentiate between risk groups.

Assessment of study quality

A validated quality assessment tool was used to evaluate the quality of the studies that met our eligibility criteria by two authors (SS and SB) [14,15]. This tool utilising a series of questions applied to each individual study to address study objectives, study population, the intervention and any co-interventions, the outcome measures, statistical analysis, results and conclusion, and disclosures.

2020 NCCN Guidelines

The NCCN has written guidelines on recommended EBRT dose and fractionation based on existing clinical evidence [16] for very low to unfavourable intermediate risk PCa. A subgroup analysis of those studies that used dosing regimens consistent with the 2020 NCCN guidelines for EBRT (including hypofractionated regimens) was performed.
Statistical analysis

Data collection and basic analysis was performed in Excel (Microsoft Corporation, Redmond, CA, USA). To determine a specific positive and indeterminate biopsy rate across a selection of studies, a weighted average of each study was performed, which combines both the reported biopsy rate and the number of patients treated:

\[
\text{Individual weights } (W_i) = N_{\text{biopsied}} \frac{\% \text{ biopsy positive}}{N_{\text{total}}}
\]

\[
\text{Weighted Average} = \sum_{i=0}^{n} W_i
\]

95% confidence intervals (CI) for the positive and indeterminate biopsy rates were calculated by obtaining the total number of patients that underwent a biopsy and the number of positive/indeterminate biopsies from each individual study. This data was entered into MedCalc (Medcalc Software Ltd, Ostend, Belgium) and using a Freeman-Tukey transformation of proportions under assumption of random effects an overall CI was derived. This technique was applied for all subgroup analysis.

For those studies where sufficient information was provided, the odds ratio (OR) of biopsy outcome on biochemical failure (BCF), distant metastases-free survival (DMFS), and prostate cancer-specific mortality (PCSM) was calculated, in a manner similar to [17]. Briefly, studies were only included if they reported both the number of positive/negative biopsies as well as the corresponding outcome between a specific positive/negative biopsy outcome and the presence/absence of BCF, DMFS and PCSM. Patients who did not undergo a biopsy, or whose biopsy finding was indeterminate, were not included in the analysis. The time to failure was not accounted for in the model.
For each eligible study a 2x2 contingency table was be computed. We calculated the total number of patients who had a positive or negative biopsy, and the proportion from each group that had BCF, DMFS and PCSM. The data was input into Cochrane Review Manager Software v5.3 (Cochrane, London, United Kingdom) and this software was used to statistically pool odds ratios (OR), CI, p-values and additionally characterize data heterogeneity. The dichotomous Mantel-Haenszel technique with a random effects model and confidence intervals of 95% were used.

RESULTS

Search results - overall

The selection process is shown in Figure 1. Twenty-two studies satisfied the eligibility criteria, including 10 randomized controlled trials and 12 cohort studies (Table 1).

Search results – post-EBRT biopsy protocol

Differences were observed in the post-EBRT biopsy protocol, which is summarized in Table 1. Post-EBRT biopsy was mandated in 68% of all studies (15/22), while patients were encouraged to undergo a post-EBRT biopsy in 27% of studies (6/22). One study did not disclose their post-EBRT biopsy methodology. Even if the post-EBRT was mandated at enrolment, many patients did not undergo their follow-up biopsy. As a result, the biopsy follow-up rates varied considerably ranging from 19% to 100% (median 59%).

Search results – PCa risk group

Seven of twenty-two studies included exclusively low- or intermediate-risk patients, with the remaining 15/22 studies including some population of high-risk patients. Of these fifteen studies including any population of high-risk patients,
only four studied predominantly high-risk patients, which accounts for 10\% of all patients included in the entire meta-analysis, and therefore the remaining 90\% of patients in the meta-analysis were either completely or predominantly low- to intermediate-risk. The risk dependent post-EBRT positive biopsy rate was directly extracted in exactly half of the twenty-two studies, while the combined positive biopsy rate across all risk groups was used for the remainder.

*Study quality evaluation*

The results of the study quality assessment are described in Table 2. Most studies met the statements of the quality assessment tool (mean 81\%, range 41 - 100\%). Some older studies did not mention competing interests or funding support (41\%). Nine studies (41\%) did not report adverse events because they focused on oncological outcomes similar to this review. Overall, according to quality assessment criteria, the quality of the studies included was high.

*Overall biopsy rates*

The weighted-average positive biopsy rate across the twenty-two studies was 32\% (95\% CI: 25-39\%, range: 4.0 - 67\%), which includes a total of 3017 patients biopsied at ≥ 2 years. This positive biopsy rate does not discriminate based on the trial design, how EBRT was delivered, whether ADT was administered, whether the radiation dose conformed to the 2020 NCCN guidelines, or how the post-EBRT biopsy protocol was defined. Across nine studies that reported it, indeterminate biopsy was identified in a weighted-average of 22\% (95\% CI: 14-28\%, range: 5.9 – 39\%) of patients.
Sensitivity analysis

Several differences across the available studies were observed. A sensitivity analysis of potentially impactful variables was performed (Table 3), along with the respective 95% CI.

Positive biopsy rate vs. NCCN 2020 guidelines, short-term ADT vs. no ADT

Nine studies used dosage regimens consistent with the current 2020 NCCN guidelines directly or had populations of patients that did (Table 3). It was possible in 8/9 studies to extract the positive biopsy rate for those patients who received dose rates that were consistent with the NCCN 2020 guidelines. One study (Nichol et al.) did not specify the positive biopsy rate as a function of dose rate. However, over 90% of patients in this specific study received a dose regimen consistent with NCCN 2020 guidelines, which was deemed acceptable. This led to a total of 832 patients with a weighted-average positive biopsy rate of 22% (95% CI: 19-41%, range: 3.6-58%).

Within the same subgroup, five of nine studies reported their post-EBRT positive biopsy rate of 34% (95% CI: 23-50%, range: 12-58%) across 349 patients without any ADT usage. On the other hand, it was found that the weighted positive biopsy rate in combination with short term (3 to 6 months) ADT was 14% (95% CI: 3.8-31%, range: 3.6-32%), across 241 patients. No information on long term ADT was clearly reported.

Positive biopsy rate vs. follow-up biopsy protocol

Fifteen of twenty-two studies mandated a post-EBRT biopsy in their protocol, resulting in a weighted positive biopsy rate of 35% (95% CI: 21-38%, range: 4.2-
58%) across 2450 patients. Within this 15-study subgroup, only 4/15 studies prospectively biopsied ≥70% of patients, and a 47% (95% CI: 5-63%, range: 4.2-58%) weighted positive biopsy rate across 798 patients was observed (Forman et al., Lukka et al., Loblaw et al., Freytag et al.). The remaining 11/15 studies had a weighted positive biopsy rate of 30% across 2279 patients (95% CI: 22-37%, range: 5.9-44%). The subgroup of six studies which did not mandate but merely encouraged patients to undergo a post-EBRT biopsy had a weighted positive biopsy rate of 29% (95% CI: 21-52%, range: 12-67%) across 438 patients.

**Positive biopsy rate vs. baseline PCa risk-group**

The risk-group dependence on positive biopsy rate was extracted in 11/22 studies for low- and intermediate-risk disease combined and 5/22 studies for high-risk disease only. Across a pool of 1567 patients, the weighted positive biopsy rate after EBRT was 25% (95% CI: 15-32%, range: 4.2 - 67%) for low- and intermediate-risk disease combined, which increased to 29% (95% CI: 20-46%, range: 12 - 67%) across a pool of 357 patients for high-risk disease.

**Risk associated with positive biopsy and BCF, DMFS and PCSM**

In this study we also assessed the relationship between positive biopsy at ≥ 2 years after EBRT with oncologically relevant longer-term outcomes of BCF, DMFS, and PCSM. From a pool of 1855 patients, those with a positive post-EBRT biopsy had approximately ten-fold higher odds of developing BCF than those with negative biopsy (OR 10.3, 95% CI: 3.7-28.7, p<0.00001), with a weighted absolute BCF rate of 67% vs. 29% for positive and negative biopsy, respectively (Figure 2a). From a pool of 1545 patients, those with a positive EBRT biopsy had approximately three times higher odds of developing distant metastasis than those with negative biopsy (OR 3.1, 95% CI: 2.1-4.7, p<0.00001), with a weighted absolute distant metastases rate of 17% vs. 5.6% for positive and negative biopsy, respectively (Figure 2b). Lastly, from a pool of 1530 patients, those with a positive-EBRT biopsy had five times higher odds of dying from their PCa than those with negative biopsy (OR 5.1, 95% CI: 2.6-10, p<0.00001), with a weighted absolute PCSM rate of 10% vs. 2.1% for positive and negative biopsy, respectively (Figure 2c). It should be noted that
there was heterogeneity observed across noted regarding the relationship between a positive biopsy and BCF, with an $I^2$ statistic of 91% from Figure 2a, and is likely a consequence of the variable positive biopsy rate observed in the studies. There was much lower heterogeneity for DM and PCSM, $I^2$ statistic 13% and 23% respectively.

DISCUSSION

This meta-analysis included twenty-two studies which reported biopsy at ≥ 2 years post-EBRT as an endpoint or study observation and utilised EBRT alone or in combination with androgen deprivation therapy (ADT) as primary treatment for low to intermediate risk PCa (PSA ≤ 20ng/ml, Gleason score ≤ 7, clinical stage ≤ T2b). The overall post-EBRT positive biopsy rate at least two years after treatment was 32%, although there was a variable biopsy rate in these studies which introduces bias. A subgroup analysis was performed to include only studies that prospectively mandated biopsy in their study protocol, resulting in a positive biopsy rate of 35% from 15 studies. Out of these studies, perhaps the most relevant results for understanding the rate of persistent local disease 2 years after EBRT, come from 4 studies that had high compliance to prospectively mandated biopsy (biopsy of ≥70% of all patients) resulting in a positive biopsy rate of 47%.

Other subgroup analyses were performed to account for modern dose regimens consistent with NCCN guidelines (positive biopsy rate 22%), baseline PCa risk group (positive biopsy rate 25% for low- to intermediate-risk only) and the addition/removal of ADT (positive biopsy rate 14% vs. 34%, ADT vs. no ADT). These findings illustrate the positive impact of modern dosing regimens, ADT and patient risk group on local disease control after EBRT.

In our analysis, 5 to 10-year follow-up data showed a positive biopsy post-EBRT was associated with higher odds of BCF, DMFS and PCSM by 10.3, 3.1 and 5.1 times, respectively. These associated poor outcomes due to failure of local tumour control are likely to represent patients with ‘radio-resistant’ tumours. These
tumours may have a different tumour biology to radio-sensitive tumours resulting in higher local recurrence and poorer outcomes. Further research is needed to characterise the tumour biology in these patients and find biomarkers which could help identify these patients early to prevent poor outcomes after EBRT.

Salvage therapies have evolved for patients with local recurrence after EBRT. Non-surgical options such as high intensity focused ultrasound (HIFU), cryotherapy and brachytherapy are now available for patients who may not be fit for salvage prostatectomy. Although no randomised controlled trial exists to compare all the different modalities, several case series suggest comparable oncological outcomes [15]. The findings of this study suggest that patients should be counselled about potential poor oncological outcomes if they have a local recurrence and adds weight to the need for active treatment of radiorecurrent prostate cancer.

There are several limitations in this meta-analysis. Perhaps most importantly, the cohort biopsy rate in the included studies was variable and not mandated in a few studies. This is a source of considerable bias in studies reporting positive biopsy rates with low cohort biopsy rates and consequently in this meta-analysis. In those studies where a biopsy was not mandatory, it is conceivable that patients with suspicious biochemical measurements were nevertheless more likely to receive one [26], which may have inflated the aggregate positive biopsy rate. Conversely, there are possible influencers which may have had the opposite effect. For instance, in certain trials patients who experienced BCF prior to the 2-year biopsy follow-up were removed from the study and not counted, likely lowering the reported positive biopsy rate [11,23,33,35]. Therefore, to discern an unbiased true positive biopsy rate, more studies are needed that mandate biopsy in their protocols and biopsy a high percentage of patients. It is clear from this study that even when biopsy is mandated prospectively in study protocols, the cohort biopsy rate is variable, likely because of its invasive nature. Although this introduces considerable bias to the analysis, it may reflect ‘real world’ clinical practice, where patients often refuse post-treatment biopsy. Post-treatment biopsy still remains contested as a measure of post-radiotherapy outcome due to its limitations of under-sampling, delayed histological resolution, and equivocal post-treatment histology [11,25,35].
Though these limitations were accounted for in the sensitivity analysis, their potential effect on positive biopsy rate cannot be discounted. Furthermore, prostate biopsy does not capture recurrences outside the gland. Since the number of studies reporting long-term oncological outcomes data was small, and there was heterogeneity between EBRT technique especially in older studies and study populations, this analysis is subject to publication and reporting bias.

In conclusion, this meta-analysis shows comparable positive biopsy rates at two years post EBRT treatment compared with other treatment modalities. A positive biopsy after radiotherapy compared to a negative biopsy has higher odds of poor long-term outcome.
CONFLICT OF INTEREST

Caroline M. Moore receives proctor fees from SonaCare Medical and UKHIFU. She has received speaker fees from Astellas and Janssen, and consultancy fees from Steba Biotech and Genomic Health.

The other authors declare no conflict of interest.

FUNDING AND ACKNOWLEDGEMENTS

This work was undertaken at the Biomedical Research Centre (BRC), University College London Hospital (UCLH), which received a proportion of the funding from the National Institute for Health Research (NIHR). The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health.

Caroline M. Moore receives research funding from the Medical Research Council, National Institute for Health Research, Cancer Research UK, Movember, Prostate Cancer UK, the European Association of Urology Research Foundation. She has research funding from Spectracure.
References


Figure 1. Flow diagram summarising selection of studies that meet inclusion criteria.
Figure 2. Odds ratio for patients with positive post-EBRT positive biopsy at ≥2 years follow-up, a) Risk of biochemical failure (BCF), b) risk of distant metastasis (DM), and c) risk of prostate cancer-specific mortality (PCSM).
Table 1. Studies meeting eligibility criteria

<table>
<thead>
<tr>
<th>Article Number</th>
<th>Author</th>
<th>Patient Risk Group</th>
<th>Technique</th>
<th>Total Dose</th>
<th>Fraction</th>
<th>EQOL-5D; &gt;10</th>
<th>Number patients in entire study with low-or intermediate-risk Pro</th>
<th>Number biopsied at ≥24 months</th>
<th>Biopsy Study protocol</th>
<th>Consistency with 2020 NCCN guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Koelting [18]</td>
<td>1980</td>
<td>4-dic/box</td>
<td>68.70</td>
<td>1.82</td>
<td>n/a</td>
<td>n/a</td>
<td>88</td>
<td>24</td>
<td>n/a</td>
</tr>
<tr>
<td>2</td>
<td>Furnas [19]</td>
<td>1993</td>
<td>4-dic/box</td>
<td>66.5 (65-68)</td>
<td>1.82</td>
<td>62-69</td>
<td>30</td>
<td>38</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>3</td>
<td>Jiang [20]</td>
<td>1995</td>
<td>4-dic/box</td>
<td>71.7 (68-74)</td>
<td>2.2-3</td>
<td>68.75</td>
<td>70</td>
<td>55</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>4</td>
<td>Lazarchuck [21]</td>
<td>1997</td>
<td>4-dic/box</td>
<td>64</td>
<td>2</td>
<td>86</td>
<td>120</td>
<td>80</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>5</td>
<td>Aminchi [22]</td>
<td>1998</td>
<td>4-dic/box</td>
<td>68.1 (64.5-72.3)</td>
<td>1.8</td>
<td>61-69</td>
<td>36</td>
<td>26</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>6</td>
<td>Cook [23]</td>
<td>2000</td>
<td>4-dic/box</td>
<td>75.6 (66-84)</td>
<td>1.8-2</td>
<td>65-66</td>
<td>496</td>
<td>136</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>7</td>
<td>Pollack [24]</td>
<td>2002</td>
<td>4-dic/box</td>
<td>78.7-70</td>
<td>2</td>
<td>78-78</td>
<td>203</td>
<td>119</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>8</td>
<td>Nikiel [25]</td>
<td>2005</td>
<td>4-dic/box</td>
<td>75.6 (67-78)</td>
<td>1.8</td>
<td>62-72</td>
<td>140</td>
<td>71</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>9</td>
<td>Lakhia [26]</td>
<td>2005</td>
<td>4-dic/box</td>
<td>52.5-60</td>
<td>2.27</td>
<td>50-60</td>
<td>50-60</td>
<td>50</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>10</td>
<td>Maria [27]</td>
<td>2007</td>
<td>4-dic/box</td>
<td>60-73.5</td>
<td>2</td>
<td>68-74</td>
<td>122</td>
<td>76</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>11</td>
<td>Zeldin [28]</td>
<td>2008</td>
<td>4-dic/box</td>
<td>70</td>
<td>2</td>
<td>70</td>
<td>120</td>
<td>55</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>12</td>
<td>Szabo [29]</td>
<td>2011</td>
<td>4-dic/box</td>
<td>52.5 (67.3-67.3)</td>
<td>2.27</td>
<td>50-60</td>
<td>50-60</td>
<td>50</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>13</td>
<td>Labba [30]</td>
<td>2013</td>
<td>4-dic/box</td>
<td>32</td>
<td>2</td>
<td>32</td>
<td>34</td>
<td>32</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>14</td>
<td>Pinto [31]</td>
<td>2013</td>
<td>4-dic/box</td>
<td>65-75</td>
<td>2</td>
<td>68-74</td>
<td>122</td>
<td>76</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>15</td>
<td>Ratto [32]</td>
<td>2014</td>
<td>4-dic/box</td>
<td>60</td>
<td>2</td>
<td>60</td>
<td>23</td>
<td>19</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>16</td>
<td>Chen [33]</td>
<td>2015</td>
<td>4-dic/box</td>
<td>68.8</td>
<td>1.8</td>
<td>68</td>
<td>170</td>
<td>83</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>17</td>
<td>Huang [34]</td>
<td>2015</td>
<td>4-dic/box</td>
<td>70-74</td>
<td>2.2-7</td>
<td>74-81</td>
<td>303</td>
<td>94</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>18</td>
<td>Kass-Hoyt [35]</td>
<td>2018</td>
<td>4-dic/box</td>
<td>75.6 (66-84)</td>
<td>1.8-2</td>
<td>78-89</td>
<td>230</td>
<td>154</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>19</td>
<td>Zeldin [36]</td>
<td>2018</td>
<td>4-dic/box</td>
<td>37.3-48</td>
<td>7.5-8</td>
<td>83-83</td>
<td>151</td>
<td>119</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>20</td>
<td>Zepnic [37]</td>
<td>2019</td>
<td>4-dic/box</td>
<td>77.6 (66-5)</td>
<td>1.8-2</td>
<td>78-89</td>
<td>232</td>
<td>232</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

3D-CRT = three-dimensional conformal radiotherapy; SBRT = stereotactic body radiotherapy; IMRT = image-guided radiotherapy; HIMRT = hypofractionated IMRT; SHRT = stereotactic hypofractionated radiotherapy

* Only a percentage of all treated patients received their EBRT treatment according to 2020 NCCN guidelines.

† Low- and intermediate-risk patients could not be separated from high-risk.
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Studies, n (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Objective</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>1. Is the hypothesis/aim/objective of the study clearly stated in the abstract, introduction, or methods section?</td>
<td>22 (100)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td><strong>Study Population</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Are the characteristics of the participants included in the study described?</td>
<td>21 (95)</td>
<td>1 (5.0)</td>
</tr>
<tr>
<td>3. Were the cases collected in more than 1 Centre?</td>
<td>9 (41)</td>
<td>13 (59)</td>
</tr>
<tr>
<td>4. Are the eligibility criteria to enter the study explicit and appropriate?</td>
<td>15 (68)</td>
<td>7 (32)</td>
</tr>
<tr>
<td>5. Did participants enter the study at a similar point in the disease?</td>
<td>21 (95)</td>
<td>1 (5.0)</td>
</tr>
<tr>
<td><strong>Intervention and co-intervention</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Was the intervention clearly described in the study?</td>
<td>21 (95)</td>
<td>1 (5.0)</td>
</tr>
<tr>
<td>7. Were additional interventions (co-interventions) clearly reported in the study?</td>
<td>20 (91)</td>
<td>2 (9.0)</td>
</tr>
<tr>
<td><strong>Outcome Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Are the outcome measures clearly defined in the introduction or methods section?</td>
<td>22 (100)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>9. Were relevant outcomes appropriately measured with objective/or subjective methods?</td>
<td>22 (100)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>10. Were outcomes measured before and after the intervention?</td>
<td>19 (86)</td>
<td>3 (14)</td>
</tr>
<tr>
<td><strong>Statistical analysis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Were the statistical tests used to assess the relevant outcomes appropriate?</td>
<td>16 (73)</td>
<td>6 (27)</td>
</tr>
<tr>
<td><strong>Results and conclusions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Was the length of follow-up reported?</td>
<td>22 (100)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>13. Was the loss of follow-up reported?</td>
<td>21 (95)</td>
<td>1 (5.0)</td>
</tr>
<tr>
<td>14. Does the study provide estimates of the random variability in the data analysis of relevant outcomes?</td>
<td>16 (73)</td>
<td>6 (27)</td>
</tr>
<tr>
<td>15. Are the adverse events reported?</td>
<td>9 (41)</td>
<td>13 (59)</td>
</tr>
<tr>
<td>16. Are the conclusions of the study supported by the results?</td>
<td>17 (77)</td>
<td>5 (23)</td>
</tr>
<tr>
<td><strong>Competing interest and source of support</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Are both competing interest and source of support for the study reported?</td>
<td>9 (41)</td>
<td>13 (59)</td>
</tr>
</tbody>
</table>

**Table 2.** Modified Delphi technique used to assess study quality on the 22 studies which met eligibility criteria

---

289
Table 3. Factors that influence post-EBRT positive biopsy rate.

<table>
<thead>
<tr>
<th></th>
<th>All studies</th>
<th>Consistent with 2020 NCCN guidelines</th>
<th>Consistent with 2020 NCCN guidelines, no ADT</th>
<th>Consistent with 2020 NCCN guidelines + short-term ADT</th>
<th>Mandated biopsy and managed to biopsy ≥70% of all patients</th>
<th>Exclusively low-intermediate-risk patients</th>
<th>Exclusively high-risk patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of studies</td>
<td>22</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>No. of patients</td>
<td>3067</td>
<td>832</td>
<td>349</td>
<td>241</td>
<td>2450</td>
<td>798</td>
<td>1567</td>
</tr>
<tr>
<td>Positive biopsy rate</td>
<td>32% (25-39)</td>
<td>22% (19-41)</td>
<td>34% (23-50)</td>
<td>14% (11-31)</td>
<td>15% (21-38)</td>
<td>47% (5-63)</td>
<td>25% (15-32)</td>
</tr>
</tbody>
</table>

Table 3. Factors that influence post-EBRT positive biopsy rate
9.1.3 PREVALENCE OF PROSTATE CALCIFICATION IN MEN WITH PROSTATE CANCER

This paper investigated the prevalence of prostate calcification in men with prostate cancer. Calcification is known to interfere with radiotherapy of the prostate and could be as a potential cause for under-treatment leading to recurrence.


Prostatic calcifications: Quantifying occurrence, radiodensity, and spatial distribution in prostate cancer patients

Background: To evaluate the prevalence, density, and distribution of prostate calcification in patients with prostate cancer.

Methods: Patients who underwent both Gallium-68 PSMA PET/CT and MRI of the prostate over the course of a year were selected for analysis. The CT images with visible calcifications within the prostate were included and calcifications automatically isolated using a threshold of 130 HU. The corresponding multiparametric MRI was assessed and the peripheral zone, transition zone, MRI-visible tumour and urethra manually contoured. The contoured MRI and CT images were registered using rigid registration, and calcifications mapped automatically to the MRI contours.

Results: A total of 85 men (age range 50-88, mean 69 years, standard deviation 7.2 years) were assessed. The mean PSA was 16.7, range 0.12 to 94.4, standard deviation 19.8. Most patients, 68%, had intermediate risk disease (Gleason grade group 2 and 3), 26% had high risk disease (Gleason grade group 4 and 5) and 6% had low risk disease (Gleason grade group 1). 46 patients out of 85 (54%) had intraprostatic calcification. Calcification occurred more in transition zone than the
peripheral zone (65% vs 35%). The mean density of the calcification was 227 HU (min 133, max 1966 HU). In 12 patients, the calcification was within an MRI-visible tumour, in 24 patients, there were calcifications within a 9 mm distance of the tumour border, and in 9 patients, there were calcifications located between the 21 urethra and tumour.

**Conclusions:** Calcifications are common in patients with prostate cancer. Their density and location may make them a significant consideration when planning treatment or re-treatment with some types of minimally invasive therapy.
INTRODUCTION

Prostatic calcifications are commonly found in men and are thought to be associated with prostatitis, chronic pelvic pain syndrome and prostate cancer [1, 2]. Previously these calcifications were not considered clinically significant, and their presence usually not mentioned in diagnostic imaging reports. Recent studies have however shown that high density material such as calcification can have a significant impact on treatment delivery in high intensity focused ultrasound (HIFU), transurethral ultrasound ablation, and brachytherapy [3-6]. In ultrasound therapy, the high density inclusions can cause reflections of the ultrasound beam, causing changes in the treated volume, which can result in over or under treatment [4, 5]. In brachytherapy, the presence of prostatic calcification changes the tissue effective atomic number, leading to altered dose distribution and potential under-dosing [3]. Improved understanding of the formation, composition, and distribution of these calcifications could allow development of treatment strategies which mitigate these effects.

The pathogenesis of prostate calcification is thought to be related to prostatic inflammation, urinary retention, or prostatic reflux [7-9]. These factors are also thought to have a role in prostate cancer and therefore it is not surprising that calcifications coexist in prostates with cancer. The source of calcification is thought to be desquamated acinar cells which form a substance called corpora amylacea. Hydroxyapatite (HA) is then deposited on corpora amylacea forming corporal calculi [7]. An alternative mechanism of HA deposition has been proposed by a group who suggest that HA is deposited by osteoblast-like epithelial cells (POLCs). The authors suggest that POLCs may be associated with prostate cancer cells and prostate calcification may be a prognostic marker [10].

A few studies have investigated the prevalence of prostate calcification using either imaging or histopathological analysis. A histological study analysed 298 consecutive whole mount prostate for patients with prostate cancer and found 88.6% contained calcifications [11]. A lower incidence of 58.8% was reported in a study of patients undergoing transrectal sonography who had prostate cancer on biopsy [12]. However, no studies have accurately mapped the location and distribution of calcifications or analysed their radiodensity using modern CT and MRI. Furthermore, there has been no investigation of calcifications in patients who
have undergone treatments such as brachytherapy, HIFU, transurethral ultrasound ablation, or cryotherapy.

The aim of this study was to accurately map and quantify prostatic calcifications using multi-modal imaging and computational tools, in a cohort of patients undergoing or having previously undergone treatment for prostate cancer.

MATERIALS AND METHODS

Ethical approval for this study was granted by the Yorkshire and the Humber Research Ethics Committee (18/YH/0411).

Study Cohort

In order to select patients who had contemporaneous CT and MRI imaging data, a consecutive cohort of patients who underwent Gallium-68 PSMA PET/CT and multiparametric MRI (within 6 months of each other) between August 2017 and August 2018 were retrospectively selected. The clinical indications for PSMA PET at our institution are mainly staging of high risk prostate cancer and assessment of cancer recurrence after treatment. Patients who had a diagnosis of prostate cancer mentioned in the clinical indications were included. Patients who had undergone radical prostatectomy were excluded. Clinical records of selected patients were reviewed for clinical data including PSA and histopathological reports.

Imaging

A total of 85 datasets were obtained from patients who had undergone both Gallium-68 PSMA PET/CT and MR scans. The whole body CT was acquired with 2.5 mm slice thickness and the modal in-plane resolution was 0.98 mm (min 0.98 mm, max 1.37 mm). The multiparametric MRI (mpMRI) included T2W small field of view images, high b-value DWI image (b1400 or b2000), ADC map and dynamic contrast enhanced images. The modal in-plane resolution for the MR images was 0.99 mm (min 0.35 mm, max 0.78 mm) and the modal slice separation was 3.3 mm (min 3 mm, max 3.85 mm). The CT images for all patients were assessed by a Board-certified Radiologist (SS).

Computational Analysis
Of these datasets, the prostate and urethra were contoured on the MR images of 45 datasets, chosen as calcifications were evident on visual inspection of the CT scans. The peripheral and transition zone base, midgland and apex regions of the prostate, as well as the urethra were contoured by a radiologist, slice by slice using Horos (horosproject.org), then exported as .xml files for further use. Signal abnormality considered as suspicious or highly suspicious for tumour (Likert score 4 or 5) was contoured if it corresponded to cancerous regions on biopsy.

Identification of calcifications was automated in order to reduce variability and increase accuracy of the output statistics. Registration of the CT and MR datasets was performed to allow translation of the contours from the MR images to the CT image space [13], where calcifications can be identified by their radiodensity. For identification of the calculi, the CT images were thresholded at 130 HU. Above this threshold, only calcified and bony tissue should be visible [14, 15] with minimal image noise on visual inspection. Each of the contours was defined as a search region, in addition to three further search regions. The first of these was derived from the tumour region, grown radially by 9 mm, which is the recommended treatment margin for focal therapy in the prostate [16]. Another region was defined as the volume located between the urethra and tumour search regions. The final region was formed from the sum of all search regions. For each of these regions, clusters of voxels with intensities above the threshold were identified computationally.

For each cluster, the coordinates of the centroid, cluster volume, principal axes lengths of an ellipsoid fitted to the cluster, and mean voxel intensities were computed. Any clusters containing a single voxel were rejected as they were usually found to have voxel intensities close to the threshold, and were indistinguishable from noise on visual inspection. The lateral (in plane) distance of the centroid of each cluster from the centroid of the urethra search region was also calculated, where the centroid of the urethra was calculated from the mean of the in plane centroid coordinates across all image slices containing the urethra contour. For each patient, mean, minimum and maximum values of these quantities were calculated for each contoured region of interest. Calcifications spanning multiple regions of interest were counted in each region. Their statistics were computed only for the part of the calcification located inside each of the zones; the statistics of
the entire calcification were computed under the total prostate region, provided it lay fully within that region.

RESULTS

Patient Cohort

A total of 85 men (age range 50-88, mean 69 years, standard deviation 7.2 years) were assessed. The mean PSA was 16.7 ng/ml, range 0.12-94.4, standard deviation 19.8. All patients had a diagnosis of prostate cancer and for 78 patients biopsy information was available in their electronic health records. In terms of Gleason grade group; 68% had intermediate risk disease (Gleason grade group 2 and 3), 26% had high risk disease (Gleason grade group 4 and 5) and 6% had low risk disease (Gleason grade group 1). Overall Gleason grade is given in Table I. 48 patients in this cohort had a history of a previous treatment. 16 patients had previous HIFU, 16 had previous external beam radiotherapy, 7 had previous brachytherapy, 3 had previous cryotherapy, 1 had previous chemotherapy and 1 had reversible electroporation. Some patients had more than one treatment type; 2 had external beam radiotherapy and salvage HIFU, 1 had brachytherapy and salvage HIFU and 1 had HIFU and cryotherapy.

Prostate Calcification

Intra-prostatic calcifications were found in 46 out of 85 patients (Table II). Examples of corresponding MR and CT images, with visible calculi within transformed region contours is shown in Fig 1. An average of 5 foci of calcifications were identified in each patient, with 24 being the highest number of calcifications identified in a single patient. All calcifications were located within 36.9 mm of the centroid of the urethra ROI, at a mean distance of 12.1 mm. Calcifications were distributed throughout the regions of the prostate, on average across 4 different regions. The mean volume of calcifications was 55.3 mm$^3$. The largest calcification, with a volume of 1263.6 mm$^3$ spanned several ROIs; it is therefore recorded under the ‘all regions’ search volume, whereas for that patient, the largest calcification found in any region individually had a volume of 689.0 mm$^3$. This represents only part of the calcification which extends beyond the boundary of this region of interest; a similar occurrence can be observed in Fig 1a. Most calcifications tended more towards ellipsoid rather than spherical in shape, with a mean aspect ratio of 1.
Results are shown in for calcifications found in the total region mask and for each ROI individually.

In 33 out of these 46 patients, a MRI visible tumour was identified. In 12 patients, there were calcifications within the tumour (see Fig 1c), and 24 patients had calcifications within a 9 mm border of the tumour (see Fig 1b).

DISCUSSION

This is the first study to accurately map prostate calcification in patients with prostate cancer using computational methods to analyse rich imaging data available from contemporaneous multi-parametric MRI and CT. Our study has shown a few important differences in the distribution of calcification compared to previous studies. Although calcification occurred more in the transition zone, there were a significant number in the peripheral zone (35%), a greater proportion than previously reported, for instance 17% \( [11] \) and 6.8% \( [12] \).

A higher incidence in the peripheral zone could be explained by differences in the patient cohort studied compared to other studies. This cohort had a high percentage of high risk disease and history of previous treatment. As the majority of prostate cancers occur in the peripheral zone and calcification can be associated with cancer, this may explain the higher incidence in our cohort. The densities of the calcifications ranged from 133 to 1966 HU, with a mean of 227 HU. This is comparable to the densities reported in previous studies conducted on smaller patient cohorts \( [3, 17] \). The volume of calcifications ranged from 4.8 to 1263.6 \( mm^3 \). The larger foci were found more in the transition zone, consistent with previous studies.

An important finding of this study is that prostatic calcification can commonly occur within and in the locale of prostate cancer. In 12 patients, foci of calcification were within MRI-visible tumours. The presence of tumour calcification in this cohort is higher than previously observed \( [11] \). This higher incidence may be explained again by many patients in this cohort having previous treatments and then having a local recurrence. We hypothesise that previous treatments such as radiotherapy or thermal ablation cause inflammation resulting in a healing response which leads to deposition of calcification. This process is seen in other organs such as the liver and kidneys \( [18, 19] \) and has been reported after
radiotherapy and thermal ablation in the prostate [20, 21]. Furthermore, we found that in 24 patients, there were calcifications within 9 mm of the tumour, which is the ablation margin recommended for focal therapy [16].

The prostatic urethra is an important structure to protect in prostate therapies. Our analysis showed that in 13 patients calcification was in a peri-urethral distribution, and in 9 patients, calcifications were located between the urethra and tumour (see Fig 1b). The presence of high density material such as calcification in the prostate has been shown to cause aberration of ultrasound waves and photon attenuation in radiotherapy. These effects can cause significant differences in treatment dose for both ultrasound ablation and radiotherapy. For transurethral ultrasound ablation, for example, calcifications located around the urethra, between the urethra and tumour, or in the tumour itself, may lie in the path of the beam. If the presence of these calcifications is not accounted for in treatment planning, then there could be a risk of under-treatment and subsequent recurrence. Further studies are needed to model the effects of prostate calcification in treatment delivery in order to determine strategies and robust thresholds for treatment.

MRI has become a key imaging modality in the assessment of prostate cancer but calcifications are difficult to visualise in standard multi-parametric protocols [22]. It is widely used to plan thermal ablation treatments such as HIFU, cryotherapy and transurethral ultrasound ablation. However calcifications are often not visible on MRI and not routinely commented on in radiological reports. Therefore, the impact of calcification in these therapies is likely under-recognised and under-reported. In contrast, for patients undergoing radiotherapy, CT imaging is usually available and calcifications can be used as pseudo-fiducial markers to plan treatment [23]. However most CT reports do not mention prostatic calcification, even though studies have shown an impact in dose delivery [3].

**CONCLUSION**

In this study, the location and density of prostatic calcifications were accurately mapped. The study has shown that prostatic calcifications are common in patients with prostate cancer. A large proportion of calcifications occur in and around tumours which could have an impact on their subsequent treatment.
CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

ACKNOWLEDGMENTS

This work was supported in part by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) (203145Z/16/Z), and by the Engineering and Physical Sciences Research Council (EPSRC).

REFERENCES


Sfanos KS, Wilson BA, De Marzo AM, Isaacs WB. Acute inflammatory proteins constitute the organic matrix of prostatic corpora amylacea and calculi in men with


TABLE I. Summary of Patient Demographics.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>85</td>
</tr>
<tr>
<td>Median Age (y)</td>
<td>70 (50-88)</td>
</tr>
<tr>
<td>Mean PSA level (ng/ml)</td>
<td>16.7 (0.12 - 94.5)</td>
</tr>
<tr>
<td>Overall Gleason Grade</td>
<td></td>
</tr>
<tr>
<td>3+3</td>
<td>5</td>
</tr>
<tr>
<td>3+4</td>
<td>28</td>
</tr>
<tr>
<td>4+3</td>
<td>25</td>
</tr>
<tr>
<td>3+5</td>
<td>1</td>
</tr>
<tr>
<td>4+4</td>
<td>5</td>
</tr>
<tr>
<td>4+5</td>
<td>13</td>
</tr>
<tr>
<td>5+4</td>
<td>1</td>
</tr>
<tr>
<td>Previous Treatment</td>
<td>48/85</td>
</tr>
</tbody>
</table>
TABLE II. Calcification statistics for all regions together and individually for each contoured region. Number of calcifications, distance from the urethra, volume and mean pixel intensity are given as mean (minimum, maximum) of the values for each patient. The distance from the urethra is defined as the in plane straight line distance between the centroid of the calcification and centroid of the urethra region, where the centroid of the urethra is calculated from the mean of the in plane centroid coordinates across all image slices containing the urethra contour. TZ = transition zone, PZ = peripheral zone.

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of patients</th>
<th>Number of calcifications</th>
<th>Urethra distance</th>
<th>Volume</th>
<th>Mean pixel intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mm</td>
<td>mm³</td>
<td>HU</td>
</tr>
<tr>
<td>All regions</td>
<td>46</td>
<td>5 (1, 24)</td>
<td>12.1 (0.3, 36.9)</td>
<td>55.3 (4.8, 1263.6)</td>
<td>227 (133, 1966)</td>
</tr>
<tr>
<td>TZ Base</td>
<td>22</td>
<td>2 (1, 12)</td>
<td>10.8 (1.4, 27.4)</td>
<td>40.4 (4.8, 522.1)</td>
<td>251 (133, 1105)</td>
</tr>
<tr>
<td>TZ Midgland</td>
<td>26</td>
<td>2 (1, 6)</td>
<td>8.6 (1.1, 28.1)</td>
<td>30.4 (4.8, 689.0)</td>
<td>227 (134, 1405)</td>
</tr>
<tr>
<td>TZ Apex</td>
<td>21</td>
<td>2 (1, 8)</td>
<td>9.1 (2.5, 25.9)</td>
<td>43.1 (4.8, 703.3)</td>
<td>230 (133, 982)</td>
</tr>
<tr>
<td>PZ Base</td>
<td>9</td>
<td>1 (1, 2)</td>
<td>13.5 (7.3, 21.6)</td>
<td>12.4 (4.8, 40.5)</td>
<td>248 (133, 755)</td>
</tr>
<tr>
<td>PZ Midgland</td>
<td>18</td>
<td>2 (1, 7)</td>
<td>14.4 (7.0, 22.7)</td>
<td>33.7 (4.8, 376.7)</td>
<td>259 (134, 1966)</td>
</tr>
<tr>
<td>PZ Apex</td>
<td>20</td>
<td>2 (1, 6)</td>
<td>11.2 (2.6, 22.9)</td>
<td>30.7 (4.8, 191.6)</td>
<td>267 (136, 1292)</td>
</tr>
<tr>
<td>Urethra</td>
<td>13</td>
<td>2 (1, 5)</td>
<td>3.6 (0.9, 5.9)</td>
<td>28.1 (4.8, 128.7)</td>
<td>212 (144, 446)</td>
</tr>
<tr>
<td>Tumour</td>
<td>12</td>
<td>1 (1, 2)</td>
<td>12.6 (4.7, 31.0)</td>
<td>62.4 (4.8, 326.6)</td>
<td>241 (138, 502)</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Tumour Margin</td>
<td>24</td>
<td>3 (1, 20)</td>
<td>13.9 (0.3, 36.9)</td>
<td>46.0 (4.8, 439.3)</td>
<td>245 (133, 690)</td>
</tr>
<tr>
<td>Between Tumour &amp; Urethra</td>
<td>9</td>
<td>1 (1, 3)</td>
<td>7.6 (4.3, 15.5)</td>
<td>54.0 (4.8, 327.1)</td>
<td>263 (153, 449)</td>
</tr>
</tbody>
</table>
FIG. 1. T2 weighted MR image slice (left), the corresponding CT image slice (middle) and thresholded image (right) for three example cases: a) previous left sided HIFU on which calcifications (highlighted by white arrows) are visible within the transition zone mid-gland contour (yellow). b) Calcification is visible between the urethra and tumour. c) Calcification is present within the tumour. The peripheral zone mid-gland (blue), urethra (purple), tumour (red), and area between urethra and tumour (purple dotted) are also shown.
Non-invasive Gleason Score Classification with VERDICT-MRI

Valindria V, Singh S, Palombo M et al. Non-invasive Gleason Score Classification with VERDICT-MRI International Society of Magnetic Resonance Medicine
March 2021

9.1.4 SELECTED CONFERENCE ABSTRACTS

Valindria V, Singh S, Palombo M et al. Non-invasive Gleason Score Classification with VERDICT-MRI International Society of Magnetic Resonance Medicine
March 2021

Non-invasive Gleason Score Classification with VERDICT-MRI

Valindria V, Singh S, Palombo M et al. Non-invasive Gleason Score Classification with VERDICT-MRI International Society of Magnetic Resonance Medicine
March 2021

Synopsis
This study proposes non-invasive Gleason Score (GS) classification for prostate cancer with VERDICT-MRI using convolutional neural networks (CNNs). We evaluate GS classification using parametric maps from the VERDICT prostate model with compensated relaxation. We classify lesions using two CNN architectures: DenseInet and SE-ResNet. Results show that VERDICT achieves high GS classification performance using SI-ResNet with all parametric maps as input. In addition to previous’s work classification multi-parametric MRI studies, VERDICT maps achieve higher metrics.

Introduction
Prostate cancer (PCa) diagnosis requires transrectal or transperineal biopsy, which is invasive and uncomfortable. Early diagnosis and treatment planning are important to reduce the mortality rate of PCa. Gleason score (GS) is the standard for measuring PCa aggressiveness and is obtained after histological inspection. Accurate GS is required to ensure prompt and correct PCa treatment.

Multi-parametric MRI (mp-MRI) is the standard care for PCa diagnosis in the UK; however, the specificity still requires improvement. Vascular Extracelular, and Restricted-diffusion for Cytometry in Tumours (VERDICT) is an advanced microscopic imaging technique for PCa characterization. VERDICT has demonstrated clear feasibility and reliability in the clinical setting with the Intracellular (IC) volume fraction (IC) demonstrating capability to discriminate GS 3+4 from GS 3+3.

Machine learning methods have been employed for GS classification using mp-MRI. A recent survey reports that machine-learning methods for GS applications using mp-MRI to solve binary classification (no cancer vs. significant cancer) rather than GS. Five-point GS classification remains a challenging task with non-invasive imaging i.e. mp-MRI with apparent diffusion coefficient (ADC), high b-value diffusion-weighted imaging (DWI), ADC images and dynamic contrast-enhanced (DCE). The best-performing methods in a prostate imaging challenge (ProstateX) achieved only a weak weighted kappa value (metric for measuring agreement between estimated GS and ground truth obtained by pathology)7. Methods using convolutional neural networks (CNNs) for GS classification with mp-MRI have not yet achieved stable results.

In this work, we evaluate GS classification using VERDICT maps generated with a recently proposed model that incorporates compartment-specific T1/T2 relaxation. We classify cancer lesions into a five-point GS using two CNNs: DenseNet12 and SE-ResNet13.

Methods

• Data Acquisition
This study is part of the INNOVATE clinical trial with 44 patients (median age, 65 years; range, 50-89 years). VERDICT-MRI acquired five b-values (90, 500, 1500, 2000, 3000 mmp2/mm2) in 3 orthogonal directions, five T1 (50/90 ms) and TR values (2482-3945 ms), voxel size 1.31.31.5 mm and matrix size 170×170×170. An experienced radiologist contoured the regions of interest (ROIs) corresponding to lesions on the VERDICT-MRI maps.

• Data Analysis

VERDICT images model: The VERDICT model with compensated relaxation is given by:

\[
S(b, TR, BR) = S_0(1 - e^{-\frac{BR}{T1}}) + \frac{D_{max}}{S_{max}}(\frac{D_{max}}{S_{max}} = \frac{8S_0^2}{\pi^2}, m, b) + \frac{D_{max}}{S_{max}}(\frac{D_{max}}{S_{max}} = \frac{2S_0^2}{\pi^2}, m, b) + \frac{D_{max}}{S_{max}}(\frac{D_{max}}{S_{max}} = \frac{\pi S_0^2}{\pi^2}, m, b)
\]

(6.1)

where func + func + func = 1. Figure 1 demonstrates the VERDICT model: vascular (VASC), intracellular (IC), extracellular-extracellular (EED) volume fraction (IC, IC, IC), cellularity, cell radius, diffusion coefficient (D), T1 relaxation, slow vascular T2 compartment (T2slow), and fast T2 by intracellular (T2ic).

Model Details: We implement Multilayer Perceptron (MLP) regression-based model training. We trained three hidden layers MLP consisting of 150 hidden units with a non-linear function rectified linear unit (ReLU) and a fully connected layer with synthetic DW-MRI signal using Eq. 1. The trained MLP then predicts and generates VERDICT maps.

Classification: We perform classification on predefined lesion ROIs with VERDICT maps using CNNs. We consider five classes: Bona (Bona), GS 3+3 (14%), GS 4+3 (20%), GS 4+2 (17%), GS 4+4 (17%), GS 4+4 (17%). For the data imbalance, we use 5 fold cross-validation (test size is 20%) with stratified randomization to preserve the percentage of samples for each class. For training, we keep the Images selecting the ROIs' bounding box and apply data augmentation using rigid transformation. We train the networks for 30 epochs with Adam optimization at a learning rate of 10^-4.

CNNs: We evaluate two CNNs. We consider DenseNet, a CNN that utilizes dense connections between layers to connect all layers. DenseNet requires less computation while maintaining high performance. The second model is SE-ResNet. SE-ResNet with embedded "Squeeze-and-Excitation (SE)" block that adaptively recalibrates feature response. SE block intrinsically introduces dynamics conditioned on the input, helping to boost feature discriminability. In SE-ResNet, SE was added before summation with the identity branch.

Baseline: We compare our results with previous studies on five-point GS classification and mp-MRI data from ProstateX-2 challenge with pre-trained InceptionV327 and VGG networks. Other studies with mp-MRI implement joint lesion segmentation and GS classification with U-Net6 and attention-based U-Net6.
Results
Fig 2. Shows the overall performance metrics of GS classification using VERDICT with DenseNet and SE-ResNet. Compared to literature, VERDICT provides much higher kappa scores. We also compare using $\delta$ and all VERDICT maps (each map is treated as channel) as input for classification, where they perform equally well for both networks. SE-ResNet with all VERDICT maps achieves the highest GS discrimination. Fig 3. Shows the GS classification using SE-ResNet and all VERDICT maps for each GS class. Fig 4. Reports the performance metrics of GS classification by SE-ResNet using as input different VERDICT maps. We can see that other maps, such as $D_{QSS}$, cellularity and $T1$ perform well.

Discussion and Conclusion
This work explores the potential of VERDICT MRI for GS classification with CNNs. Our results show that VERDICT with two different CNNs yields a high kappa score, substantially improvement (44%) over those found in literature with mp-MRI. Future work will include direct mp-MRI comparison with VERDICT on the same patients and a larger cohort.

Acknowledgements
This project received funding from Engineering and Physical Sciences Research Council grant (EPSRC EP/N021967/1: INNOVATE trial). UKRI Future Leaders Fellowship (MRT002096/1) supports MP, the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre.

References

Figures

Figure 1. Flowchart of Gleason score classification. We classify the predefined lesion ROIs on the VERDICT maps using DenseNet and SE-ResNet. The network then gives the corresponding Gleason score to the lesion.
Figure 2. Results of five-point GS classification using two different networks: DenseNet and SE-ResNet on VERDICT. SE-ResNet with VERDICT generally yields better performance than DenseNet. VERDICT gives higher classification metrics compared to those reported in GS classification with bi- and multi-parametric MRI.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No cancer</td>
<td>0.818</td>
<td>0.900</td>
<td>0.857</td>
</tr>
<tr>
<td>3+3</td>
<td>1.000</td>
<td>0.600</td>
<td>0.750</td>
</tr>
<tr>
<td>3+4</td>
<td>0.800</td>
<td>1.000</td>
<td>0.889</td>
</tr>
<tr>
<td>4+3</td>
<td>0.893</td>
<td>0.800</td>
<td>0.842</td>
</tr>
<tr>
<td>&gt;4+3</td>
<td>1.000</td>
<td>0.600</td>
<td>0.750</td>
</tr>
</tbody>
</table>

Figure 3. Classification report (precision, recall, and F1-score) from GS classification using all VERDICT maps with SE-ResNet. GS 3+4 gives the highest metrics, while GS 4+3 and GS >4+3 are the hardest to classify.

<table>
<thead>
<tr>
<th>VERDICT maps</th>
<th>Accuracy</th>
<th>Kappa</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWLC</td>
<td>0.815</td>
<td>0.716</td>
<td>0.811</td>
</tr>
<tr>
<td>RCC</td>
<td>0.846</td>
<td>0.773</td>
<td>0.830</td>
</tr>
<tr>
<td>FFEF</td>
<td>0.831</td>
<td>0.809</td>
<td>0.832</td>
</tr>
<tr>
<td>R</td>
<td>0.800</td>
<td>0.868</td>
<td>0.802</td>
</tr>
<tr>
<td>Cellularity</td>
<td>0.846</td>
<td>0.760</td>
<td>0.846</td>
</tr>
<tr>
<td>DEE</td>
<td>0.846</td>
<td>0.876</td>
<td>0.848</td>
</tr>
<tr>
<td>Thaw/Freeze</td>
<td>0.800</td>
<td>0.700</td>
<td>0.797</td>
</tr>
<tr>
<td>T2c</td>
<td>0.815</td>
<td>0.885</td>
<td>0.817</td>
</tr>
<tr>
<td>T1</td>
<td>0.874</td>
<td>0.837</td>
<td>0.874</td>
</tr>
</tbody>
</table>

Figure 4. Evaluation metrics for GS classification on different input of VERDICT maps individually with SE-ResNet. In addition to $C_p$ map, T1, cellularity and $D_{net}$ maps also have high performance and may help the GS classification.