

Focused Diffusion-Weighted Imaging in Prostate Cancer

FODIP

Dr Keith Craig Godley

Doctor of Medicine

University of East Anglia

Norwich Medical School

September 2017

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Background: MRI of the prostate is used to detect and localise prostate cancer and diffusion-weighted imaging (DWI) is a routine sequence. The purpose of the study was to determine the diagnostic accuracy of high b-value DWI through meta-analyses. Additionally, the aim was to determine the diagnostic accuracy of a novel combined small field-of-view (FOV) and high b-value DWI sequence for detecting and localising prostate cancer.

Material and Methods: Meta-analyses of diagnostic test accuracy of high b-value DWI, and T2WI and DWI combined, were performed conforming to the PRISMA statement. A prospective observational diagnostic test accuracy study of 40 patients who underwent 3T MRI with small-(sDWI) and conventional (cDWI) FOV and subsequent radical prostatectomy were included. Two blinded readers assessed the lesion and sectoral-based accuracy of both sequences using radical prostatectomy (RP) specimens as reference standard. Apparent Diffusion Coefficient (ADC) of benign and malignant tissue was assessed.

Results: The sensitivity, specificity and area-under-curve (AUC) of the high b-value DWI, and T2WI and DWI meta-analyses were 0.59, 0.92, and 0.92; and 0.68, 0.84 and 0.83, respectively. There were 83 prostate cancers detected in the RP specimens and half of prostate sectors contained tumour. Sensitivity of cDWI alone was significantly higher than sDWI (0.62 vs 0.45, $p < 0.001$), but specificity was lower (0.88 vs 0.92, $p < 0.001$). The AUC was not significantly different (0.76 vs 0.73, $p = 0.164$). The interobserver reliability of cDWI was substantial ($\kappa = 0.63$) and sDWI was moderate ($\kappa = 0.59$). The median mean ADC of tumours was significantly lower than non-tumour tissue for both sequences ($p < 0.001$). The ADC of tumour was significantly lower with sDWI ($960 \times 10^{-6} \text{ mm}^2/\text{sec}$ vs $766 \times 10^{-6} \text{ mm}^2/\text{sec}$, $p < 0.001$).

Conclusion: Conventional DWI outperformed small-FOV DWI imaging at lesion detection, but there may be clinical benefit for small-FOV at excluding tumour, particularly in a pre-biopsy population.

Table of Contents

Abstract	2
Table of Contents	3
List of tables	7
List of Figures	9
Acknowledgements	14
Acknowledgements specific to aspects of the thesis	15
Declaration	15
Word Count	15
Chapter 1 Introduction	16
The prostate gland	16
Prostate cancer histology and grading	19
Prostate cancer grading.....	21
Prostate cancer staging	25
Biochemical and pathological investigation of prostate cancer	27
Treatment of localized prostate cancer.....	30
Multiparametric MRI	33
DWI in prostate MRI	42
Diagnostic test accuracy studies	50
Chapter 2 Meta-analysis: Accuracy of high b-value DWI	55
Background	55
Material and Methods.....	59
Results	63
Discussion	76

Chapter 3 Meta-analysis: Accuracy of DWI and T2WI.....	82
Background	82
Materials and methods.....	83
Results	87
Discussion	101
Summary of meta-analyses and aims of FODIP study	107
Chapter 4 Materials and Methods.....	109
Summary.....	109
Research questions	109
Routine clinical care of patients with prostate cancer at NNUH	110
Study procedure	112
MRI assessment	114
Radiologist assessment.....	118
Correlation of MRI and Histopathology.....	128
Statistical Analysis	138
Patient Confidentiality.....	140
Ethical Considerations	140
Chapter 5 Results	144
Patient descriptive statistics	144
Histopathology descriptive statistics.....	146
Diagnostic performance of DWI sequences: sector-based analysis.....	155
Sectoral reliability.....	163

Diagnostic performance of DWI sequences: lesion-based analysis	164
Quantitative analysis of lesions.....	171
Correlation of tumour ADC value and tumour aggressiveness.....	188
Chapter 6 Discussion	191
Accuracy of the DWI sequences.....	191
Reliability of DWI sequences and comparison to current literature.....	201
Small-field of view results compared to current literature	204
Accuracy results of DWI sequences compared to the meta-analyses and current literature	205
Comparison of b1000 and b1500 results.....	210
ADC value of tumours.....	211
Correlation of Gleason Score and tumour ADC value	212
Influence of post-biopsy haemorrhage on results	215
Further limitations.....	218
Wider implication of results and potential future research	221
Chapter 7 Summary of FODIP study	224
Definitions.....	227
References	228
Appendix 1	252
Published material and presentations.....	252
List of formal teaching and courses attended	260
Appendix 2	262
Radiologist reporting instructions	262

Histopathology reporting instructions268

Appendix 3271

Favourable opinion of proportionate review at Research and Ethics Committee subject to conditions – 18th January 2016271

Favourable opinion of proportionate review at Research and Ethics Committee – 22nd January 2016275

Non-substantial amendment approval – 15th August 2016.....277

List of tables

Table 1-1. D'Amico prostate cancer risk groups.	24
Table 1-2. AJCC Prostate cancer TNM staging.	26
Table 2-1. MEDLINE search strategy	59
Table 2-2. Principle characteristics of included studies.	65
Table 2-3. Imaging and methodological characteristics of the included studies.	66
Table 2-4. Description of threshold method used for included studies using ADC measurements.	68
Table 2-5. Description of threshold method used for included studies using visual assessment of tumour.	68
Table 2-6. Quality assessment of the included studies.	69
Table 2-7. Diagnostic performance of eligible studies and subsets.	70
Table 2-8. Results of the subgroup analysis.	75
Table 3-1. MEDLINE search terms and strategy.	84
Table 3-2. Principle characteristics of included studies.	90
Table 3-3. Imaging and methodological characteristic of included studies.	91
Table 3-4. Diagnostic Performance of included studies.	92
Table 3-5. Sub-group analysis and meta-regression.	94
Table 4-1. Imaging parameters of sequences relevant to the FODIP study.	114
Table 5-1. Reader 1 frequency of Likert scores in different protocols.	155
Table 5-2. Reader 2 frequency of Likert scores in different protocols.	155
Table 5-3. Diagnostic performance of reader 1 for each imaging protocol on a per sector basis.	158
Table 5-4. Diagnostic performance of reader 2 for each imaging protocol on a per sector basis.	158
Table 5-5. Diagnostic performance of reader 1 and comparison of cDWI and sDWI sequences.	160
Table 5-6. Diagnostic performance of reader 2 and comparison of cDWI and sDWI sequences.	161
Table 5-7. Interobserver reliability, as demonstrated by weighted kappa tests, between reader 1 and 2 for all sectors, and the peripheral zone and transitional zone sectors.	163
Table 5-8. Index lesion identification of reader 1 and 2 with all imaging protocol following a stringent and clinical assessment of readers' results.	164

Table 5-9. Reader 1 lesion accuracy.....	165
Table 5-10. Reader 2 lesion accuracy.....	166
Table 5-11. Table of signal intensities of normal and tumour tissues in cDWI sequences.	171
Table 5-12. Table of signal intensities of normal and tumour tissues in sDWI sequences.	172
Table 5-13. Median noise measurements of sDWI and cDWI sequences.....	174
Table 5-14. Tumour SNR, CNR and CR on cDWI and sDWI sequences and comparison between DWI sequences.	175
Table 5-15. Comparison quantitative imaging characteristics of index and non-index tumours on cDWI sequence.	182
Table 5-16. Comparison quantitative imaging characteristics of index and non-index tumours on sDWI sequence.	183
Table 5-17. Quantitative imaging parameters and comparison of tumours situated within and outwith haemorrhage on cDWI sequence.	185
Table 5-18. Quantitative imaging parameters and comparison of tumours situated within and outwith areas of haemorrhage on sDWI sequence.	186
Table 5-19. Mean and minimum ADC values of tumours of different Gleason Grade Group and for tumours of Gleason Grade 3+4 and those above.....	188

List of Figures

Figure 1-1. Diagram of the prostate demonstrating the different glandular and stromal compartments, and the urethra, ejaculatory ducts and seminal vesicles.	18
Figure 1-2. T1WI of the prostate demonstrating the homogenous low signal of the gland and lack of differentiation of PZ and TZ.....	34
Figure 1-3. T2WI of a normal prostate demonstrating the anatomical divisions of the prostate and seminal vesicles. A - Axial slice through the base of the gland. B - Axial through the mid-gland. C - Axial through the apex. D - Coronal slice.	38
Figure 1-4. Diffusion of water molecules. A - Highly cellular environment in which the diffusion of water is restricted due to impedance by the extracellular space and cell membranes. B - Free movement of water in large extracellular space or due to faulty cell membranes.	39
Figure 1-5. Water diffusion metrics demonstrating the attenuation signal in different tissues between the diffusion sensitising gradients with static molecules, such as in tumour, maintaining signal and free-moving molecules, such as the bladder, losing signal.	40
Figure 2-1. Flow diagram of study identification and exclusion.....	64
Figure 2-2. QUADAS-2 results summarising the proportion of low, high and unclear risk of bias and applicability concerns.....	69
Figure 2-3. Forest plot of sensitivity with pooled sensitivity, Q statistic of the chi-squared, and I-squared results.	71
Figure 2-4. Forest plot of specificity with pooled specificity, Q statistic of the chi-squared, and I-squared results.	72
Figure 2-5. Forest plot of positive likelihood ratio with pooled positive likelihood ratio, Q statistic of the chi-squared, and I-squared results.	72
Figure 2-6. Forest plot of negative likelihood ratio with pooled negative likelihood ratio, Q statistic of the chi-squared, and I-squared results.	73
Figure 2-7. The summary Receiver Operating Characteristic (sROC) curve for high b-value DWI in detecting prostate cancer.	73
Figure 2-8. Sensitivity and 1-specificity plotted in Receiver Operating Characteristic curve for individual studies and subsets.....	74
Figure 3-1. PRISMA Flow Diagram.	88
Figure 3-2. QUADAS-2 results summarising the proportion of low, high and unclear risk of bias and applicability concerns.....	89
Figure 3-3. Forest plot of sensitivity for detecting prostate cancer including 95% CI, I ² value and Q statistic.	95

Figure 3-4. Forest plot of specificity for detecting prostate cancer including 95% CI, I ² value and Q statistic.	96
Figure 3-5. Forest plot of positive likelihood ratio for detecting prostate cancer including 95% CI, I ² value and Q statistic.	97
Figure 3-6. Forest plot of negative likelihood ratio for detecting prostate cancer including 95% CI, I ² value and Q statistic.	98
Figure 3-7. Summary receiver operating characteristic (SROC) curve for the detection of prostate cancer.	99
Figure 3-8. Sensitivity and 1-specificity plotted in Receiver Operating Characteristic curve for individual studies and subsets.	100
Figure 4-1. T1 weighted image of the prostate demonstrating post-biopsy haemorrhage in the peripheral zone bilaterally (red arrows).	116
Figure 4-2. Histopathology and imaging assessment timeline.	116
Figure 4-3. Example of malignant lesions on T2WI axial images. a) A focus of Gleason Score 3+4=7 prostate cancer in the left peripheral zone at 5 o'clock. b) A focus of Gleason Score 4+3=7 prostate cancer in the transitional zone anteriorly (red arrow).	117
Figure 4-4. Example of the same left peripheral zone prostate tumour on different DWI sequences and their respective ADC maps. On the DWI images the tumour appears as high signal relative to surrounding non-tumour tissue and on ADC maps the tumour is lower signal than non-tumour tissue. The same Gleason Score 4+3=7 tumour is demonstrated on a) b1500 cDWI. b)ADC map cDWI. c) b1500 sDWI. d) ADC map sDWI.	118
Figure 4-5. Prostate reporting diagram divided into 12 sections.	120
Figure 4-6. Radiology reporting template.	121
Figure 4-7. Completed radiologist template for case 16.	122
Figure 4-8. Histopathologist reporting template.	124
Figure 4-9. Completed histopathologist template for case 3.	126
Figure 4-10. Case 3 histopathology slides.	127
Figure 4-11. Left transitional zone prostate cancer with ROI drawn around the lesion on all three b-value images and ADC map. The corresponding T2WI (top right) and T1WI weighted slice are also demonstrated.	133
Figure 4-12. sDWI, T2WI and T1WI sequences of case 7. ROI drawn on normal peripheral and transitional zones, and on the left obturator internus muscle.	135
Figure 5-1. Distribution of ages of included patients.	144
Figure 5-2. Histogram of time from biopsy to MRI.	145
Figure 5-3. Biopsy Gleason Grade Group distribution.	146

Figure 5-4. Frequencies of tumour Gleason Grade in the included tumours.....	147
Figure 5-5. Frequencies of tumour Gleason Grade Group in the included tumours.	147
Figure 5-6. Bar chart demonstrating the number of tumours which extended beyond the prostate capsule within different Gleason Grade Groups.	148
Figure 5-7. Boxplot of maximum axial diameter of different tumour Gleason Grade Groups.	149
Figure 5-8. Boxplot of tumour volume of different tumour Gleason Grade Groups.....	150
Figure 5-9. Comparison of tumour diameter of index and non-index lesions.	151
Figure 5-10. Comparison of tumour volume of index and non-index lesions.	151
Figure 5-11. Comparison of tumour diameter of tumours confined to and those which extended beyond the capsule.....	152
Figure 5-12. Boxplot of tumour volume of tumours which extended beyond the capsule and those which did not.	153
Figure 5-13. Boxplot displaying the difference in maximum axial diameters of tumours which did and did not invade the seminal vesicles.	154
Figure 5-14. Boxplot of tumour volume of lesions which did and did not invade the seminal vesicles.	154
Figure 5-15. Bar chart demonstrating the ratio of sectors with tumour present or absent to probability of the tumour presence with the conventional DWI sequence. A – Reader 1 Protocol A, B – Reader 1 Protocol B, C – Reader 2 Protocol A, D - Reader 2 Protocol B.	156
Figure 5-16. Bar chart demonstrating the ratio of sectors with tumour present or absent to probability of the tumour presence with the small FOV DWI sequence. A - Reader 1 Protocol A, B - Reader 1 Protocol B, C - Reader 2 Protocol A, D - Reader 2 Protocol B.....	157
Figure 5-17. Receiver operating characteristic curve of sectoral diagnostic performance of reader 1.	162
Figure 5-18. Receiver operating characteristic curve of sectoral diagnostic performance of reader 2.	162
Figure 5-19. Influence of peri-lesional haemorrhage on Radiologist 1's ability to detect and localise tumour with cDWI and T2WI.....	167
Figure 5-20. Influence of peri-lesional haemorrhage on Radiologist 1's ability to detect and localise tumour with sDWI and T2WI.....	167
Figure 5-21. Influence of peri-lesional haemorrhage on Radiologist 2's ability to detect and localise tumour with cDWI and T2WI.....	168
Figure 5-22. Influence of peri-lesional haemorrhage on Radiologist 2's ability to detect and localise tumour with sDWI and T2WI.....	168

Figure 5-23. Comparison of detection rate of tumours of different Gleason Grade Group with cDWI by Radiologist 1.	169
Figure 5-24. Comparison of detection rate of tumours of different Gleason Grade Group with sDWI by Radiologist 1.	169
Figure 5-25. Comparison of detection rate of tumours of different Gleason Grade Group with cDWI by Radiologist 1	170
Figure 5-26. Comparison of detection rate of tumours of different Gleason Grade Group with sDWI by Radiologist 2.	170
Figure 5-27. Histograms of selected signal intensities demonstrating parametric and non-parametric distributions and their Shapiro-Wilk test p-value. A - Mean ADC value of tumour, B - Minimum ADC value of tumour, C - b1000 and D – b1500 signal intensity of normal tissue.	173
Figure 5-28. Histogram of noise level of A - sDWI ADC map and B – cDWI ADC maps. The p-value represents the Shapiro-Wilk test indicating a non-parametric distribution.....	174
Figure 5-29. Histogram of selected tumour quantitative results of different sequences with Shapiro-Wilk test p-values demonstrated. A - cDWI ADC SNR, B - sDWI ADC SNR, C - cDWI b1000 CR, D - sDWI b1000 CR, E - cDWI b1500 CNR, F - sDWI b1500 CNR.	176
Figure 5-30. Boxplot of mean tumour ADC values of sDWI and cDWI ADC maps.....	177
Figure 5-31. Boxplot of minimum tumour ADC values of sDWI and cDWI ADC maps.....	178
Figure 5-32. Comparison of SNR of tumour on b1000 and b1500 images using a Wilcoxon Signed Rank test. A - cDWI, B – sDWI.	179
Figure 5-33. Comparison of CNR of tumour on b1000 and b1500 images using a Wilcoxon Signed Rank test. A - cDWI, B – sDWI.....	179
Figure 5-34. Comparison of CR of tumour on b1000 and b1500 images. A - cDWI (paired samples t-test), B - sDWI (Wilcoxon Signed Rank test).....	180
Figure 5-35. Receiver operating characteristic curve of mean ADC values of the cDWI sequence.	181
Figure 5-36. Receiver operating characteristic curve of mean ADC values of the sDWI sequence.	181
Figure 5-37. ROC curve of mean ADC value of index lesions on each ADC map.	184
Figure 5-38. ROC curve of mean ADC value of sDWI and cDWI sequences including only tumours not affected by haemorrhage.....	187
Figure 5-39. Boxplot and Spearman Rank correlation coefficient of ADC of different tumour grades on cDWI. A - Mean tumour ADC, B - Minimum tumour ADC.....	189
Figure 5-40. Boxplot and Spearman Rank correlation coefficient of ADC of different tumour grades on sDWI. A - Mean tumour ADC, B - Minimum tumour ADC.	189

- Figure 5-41. Boxplot and Spearman Rank correlation coefficient of tumour ADC of Gleason Score 3+4=7 lesions and those greater than Gleason Score 3+4=7 on cDWI. A - Mean tumour ADC, B - Minimum tumour ADC.....190
- Figure 5-42. Boxplot and Spearman Rank correlation coefficient of tumour ADC of Gleason Score 3+4=7 lesions and those greater than Gleason Score 3+4=7 on sDWI. A - Mean tumour ADC, B - Minimum tumour ADC.....190
- Figure 6-1. Multifocal prostate cancer. A - cDWI b1500, B - cDWI ADC map, C - sDWI b1500, D - sDWI ADCmap, E - Pathology template report, F - Image of pathology slide.192
- Figure 6-2. Example of a tumour missed on initial assessment by the histopathologists. A - cDWI b1500, B - cDWI ADC map, C - sDWI b1500, D - sDWI ADC map.....199
- Figure 6-3. Diagram based on the methodology by Isebaert et al. (2012) [141] demonstrating the correlation of the different segments of the prostate on histopathology and radiology.....201
- Figure 6-4. Example of peripheral zone tumours. The left PZ tumour was detected. The right PZ tumour in an area of haemorrhage was not seen. A) cDWI ADC map, B) cDWI b1500, C) sDWI ADC map, D) cDWI b1500, E) T1WI , F) Correlate pathology slide.216

Acknowledgements

The execution of this project and writing of this thesis was one of the most challenging undertakings I have faced. Without a loving family, supportive colleagues, and encouraging supervisors I would surely have faltered.

Firstly, I would like to thank my wife, Lou. Your unwavering love, patience and emotional support was astonishing. You have taught me so much about sacrifice and compromise. You never fail to provide the right help at the right time. I hope I return to 'normal'.

To my son, Arthur. You did your best to derail the project by giving me an extremely tempting excuse to play with you and I thank you for being such a great sleeper.

I am forever indebted and thankful to my parents for their love, support and encouragement. I owe so much to you both.

To my supervisors, Professor Andoni Toms and Dr Paul Malcolm, I thank you for your relentless support, guidance and encouragement, particularly when I was even more dour and bleak than normal. This was only made possible by your clinical and research experience, and willingness to share it.

Additionally, I am grateful for the input of Professor Glyn Johnson and Dr Tristan Barrett, whose invaluable knowledge of MRI physics and MRI of the prostate, respectively, was of great benefit.

To the staff of the Radiology Department and Norwich Radiology Academy I am extremely appreciative for your moral support and assistance.

Acknowledgements specific to aspects of the thesis

Tom Syer was a great help in both meta-analyses. Under the guidance of Dr Toby Smith, we both performed the searches, quality assessment, data extraction and analysis of the meta-analyses. I took the lead on data analysis and the results in the high b-value meta-analysis (chapter 2), with Tom reciprocating in the DWI and T2WI meta-analysis (chapter 3). Tom also contributed to the FODIP study proper, patiently and meticulously collecting quantitative imaging data.

Dr Paul Malcolm and Dr Tristan Barrett, radiologists, performed the image assessment and completion of the reporting templates for the cases in the FODIP study.

Professor Richard Ball and Dr Ryan Butel, histopathologists, diligently and meticulously assessed the radical prostatectomy specimens. Dr Butel was a specific help in the review of challenging cases and queries regarding the pathological process.

Declaration

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Keith Godley

Word Count

58541 words (excluding appendices).

Chapter 1 Introduction

The prostate gland

Anatomy of the prostate gland

The prostate is an inverted cone shape with its base lying cranially and apex caudally. It lies posterior to the pubic symphysis, anterior to the rectum and proximal anus, and its base lies against the neck of the bladder. The prostatic urethra passes through the gland, as do the ejaculatory ducts. It lies in the subperitoneal compartment between the peritoneal cavity and pelvic diaphragm.

In the adult male, without significant hyperplasia, it weighs approximately 20-30g. It is the largest accessory gland in the body and secretes a fluid which forms part of the seminal fluid. It is also a composite gland and is composed of three glandular components: the peripheral zone (PZ), central zone (CZ), and transitional zone (TZ). It has a stromal component known as the anterior fibromuscular stroma (AFMS). The gland can also be divided into regions anatomically delineated by dividing the gland in thirds in the height dimension. These are known as apex, mid-gland, which contains the verumontanum in the mid-prostatic urethra, and base (from caudal to cranial).

The PZ is the largest of the zones, comprising 70% of the glandular tissue, wraps around the posterolateral aspect of the gland, and contains the distal prostatic urethra. It is present in the apex, mid-gland and base of the prostate and is almost a complete ring, prevented from being so by the AFMS [1,2] (Figure 1-1).

The CZ is comprised of a pair of conical structures with the apex in the mid-gland at the verumontanum and extends superolaterally to the prostate base adjacent to the seminal vesicles. They contain the ejaculatory ducts and in total account for 25% of the glandular tissue [3].

The TZ is the smallest of the glandular components in a normal prostate, comprising of 5% of normal glandular tissue, is present in the base and midgland, and is a pair of globular structures. This zone typically enlarges in older men due to benign prostatic hyperplasia (BPH).

The AFMS is comprised of smooth muscle and fibrous tissue and lies on the anterior surface of the prostate in the midline and extends from base to apex.

The prostate is not surrounded by a true capsule, but is covered incompletely by two layers; an inner layer of smooth muscle and an outer layer of collagen [4]. The degree of coverage is variable [5], but is deficient anteriorly at the AFMS, at the extreme apex anteriorly, and at the base [6,7]. Although, not technically a true capsule and inseparable from the prostate at surgery it is common to use the term capsule when reporting magnetic resonance imaging (MRI) of the prostate. Detection of extension of tumour beyond the capsule is of prognostic importance and can alter management of prostate cancer.

The seminal vesicles are a pair of grape-like structures which lie between bladder and rectum and extend superolaterally from the prostate base. The base of the seminal vesicle is at its caudal aspect lateral to the vas deferens. The duct of the seminal vesicle joins with the vas deferens to form the ejaculatory duct which pass through the CZ to the verumontanum and prostatic urethra. The size and appearance of the seminal vesicles can be variable and depends on age and sexual activity [8].

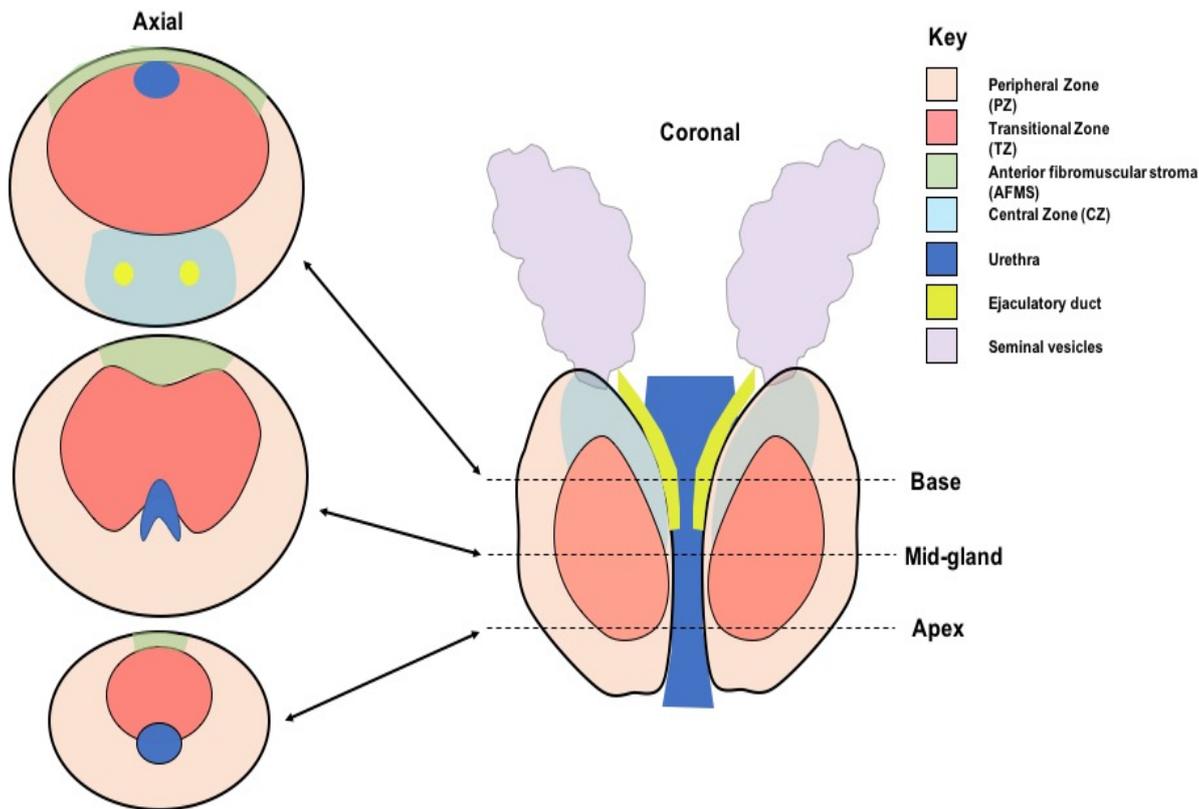


Figure 1-1. Diagram of the prostate demonstrating the different glandular and stromal compartments, and the urethra, ejaculatory ducts and seminal vesicles.

An understanding of the prostatic zonal anatomy is important when interpreting and reporting MRI. Increasingly clinicians require the information on the location of tumour within the gland to guide biopsy and target in focal therapies.

Prostate cancer histology and grading

Epidemiology of prostate cancer

Prostate cancer is the most common cancer in men, making up 26% of cancers in men in England and Wales in 2010 [9]. There was a rapid rise in incidence between 1998 and 2001 which was thought to be due to the increased use of the Prostatic Specific Antigen (PSA) blood test. In 2014 there were 46,490 new cases of prostate cancer in the UK and over the last decade the incidence rate has increased by 6% [10]. The number of new diagnoses in England and Wales is highest in those aged between 65-79 years, but the greatest increase in new diagnoses has been in the 45-59 year age group.

Prostate cancer is the 2nd highest cancer killer in men after lung cancer [11]. In 2010 the age standardized mortality rate for prostate cancer was 23.8 per 100,000 [9]. There were 11,287 deaths from prostate cancer in the UK in 2014, but the mortality rate has been in decline since 2001 [12]. This can be attributed to the increase in the use of PSA test which lead to increased cases diagnosed and also survival through lead time bias [13].

With the prevalence high and incidence rising the investigation and management of prostate cancer is a large expense. The spectrum of disease burden for patients is wide, from those with aggressive disease requiring active treatment to those with indolent disease which will never cause morbidity or mortality. With more men being diagnosed with prostate cancer it is important to have diagnostic tools which can easily and reliably risk stratify patients.

There are three well established risk factors for developing prostate cancer: increasing age, ethnicity, and hereditary factors. Only 1% of prostate cancers are found in those less than 50 years, approximately one-third are diagnosed over the age of 70 years and the remainder between these age groups [14]. This is partly due to the increased use of PSA testing in the 50 - 80-year group, with decreased use of the test in those greater than 80 years because there are

decreased benefits in screening asymptomatic men at this age. In the UK black men are three times more likely to be diagnosed with prostate cancer than white men, a difference which is greater amongst younger men [15]. In addition, this study found that men of African and Carribean origin had similar risks indicating a possible common genetic aetiology.

Approximately 1 in 4 prostate cancers occur in family clusters and about 1 in 10 are hereditary. Men with a first degree relative with the disease have a 2-3 fold increase in developing prostate cancer compared to those without [16]. There have been over 100 gene mutations associated with increased familial risk, including BRCA 1 and BRCA 2, with mutations in these genes associated with more aggressive disease and poorer outcomes [17,18].

Androgens play an important role in the growth of normal prostate and in prostate cancer, however, it's effect at the initial development of prostate cancer is unclear as men on androgen replacement therapy are not at increased risk. Furthermore, drugs which inhibit the androgen receptors have been investigated as a potential means of preventing prostate cancer, but thus far they have only been proven to decrease the incidence of low grade cancers and potentially increase the incidence of higher grade disease [19].

Prostate cancer histology

Cancer of the prostate is most typically usual-acinar adenocarcinoma with multiple but infrequent variant forms of acinar and non-acinar subtypes outlined by the WHO. The non-acinar variants or types account for 5–10% of prostate cancers, of which there are 7, with ductal adenocarcinoma the most common, accounting for 3% of prostate cancers [20]. More recently several further cancer subtypes have been described due to their specific clinico-pathological features [21]. Intraductal adenocarcinoma of the prostate is an aggressive subtype and is associated with a higher grade, larger tumours and higher probability of extracapsular extension and seminal vesicle

invasion [22]. The individual sub-types of prostate cancer are important prognostically, but are of less importance radiologically.

Prostate cancer grading

Gleason Grade

The principle features of Donald Gleason's five grades of prostate cancer, outlined first in 1966 [23], has stood the test of time and is still used today due to its reliability and simplicity. Gleason Grade 1 is considered the most differentiated and grade 5 the least. It is an important factor in determining treatment, predicting biochemical failure and the likelihood of nodal and distant metastasis after therapy, and providing prognostic and mortality information [24]. Each tumour focus is provided a Gleason Score, which is usually sum of the most and second most prevalent Gleason Grade patterns and ranges from 2-10 [25].

The 2005 ISUP modified Gleason grading system

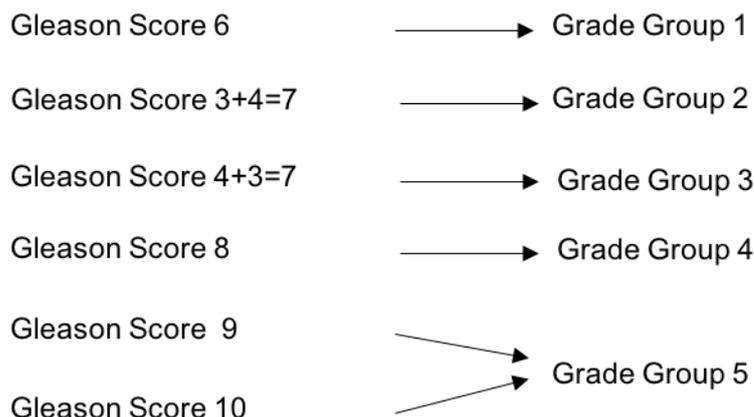
In 2005 the International Society of Urological Pathologists (ISUP) consensus on prostate cancer grading made some changes to the Gleason grading system which have been universally used since and have had a significant impact on prostate cancer investigation and management [26]. For example, they decided that prostate cancer with grade of 2 or less and a score of less than 6 should not be reported with Gleason Grade starting at 3 and Gleason Score at 6. Those tumours with a secondary pattern which is higher grade than the primary pattern should always be reported, whilst tumours with a secondary pattern of lower grade than the primary should only be reported if the secondary pattern is present in over 5% of the tumour focus. In addition, the consensus decided to change the way tertiary patterns are decided on biopsy specimens. When a high-grade tertiary pattern (such as grade 4 or 5) is identified the primary pattern and the high-grade tertiary pattern should be used to make the Gleason score.

The results of the changes at the ISUP consensus meeting was an increase in a significant percentage of previously Gleason Score 6 tumours being re-assigned a score of 7 due to tertiary pattern 4 areas. Multiple studies have assessed the changes with an increase in Gleason Score 7 disease between 13-42% and decrease in Gleason Score 6 tumours between 19–26% [27,28]. The result has been a more homogenous group of Gleason Score 6 patients with a uniformly excellent prognosis. Eggener et al. demonstrated that only 0.03% of patients with organ-confined Gleason Score 6 disease following prostatectomy died of prostate cancer using the modified scoring system [29]. The difficulty comes in confidently predicting such a favourable prognosis following a Gleason Score 6 on biopsy. Pierorazio et al. demonstrated almost one third of Gleason Score 6 tumours diagnosed on biopsy were upgraded to a score of 7 or more following radical prostatectomy [30]. However, when assessing the 5 year biochemical recurrence free survival of patients with a score of 6 on biopsy was 94.7% compared to 82.7% for those with a score of 7 suggesting a significant prognostic value in diagnosing Gleason score 6 on biopsy.

The influence of the modified Gleason scoring system has resulted in improved concordance between biopsy specimens and radical prostatectomy specimens by approximately 15% [27,31]. Biopsy undergrading is the most common cause of discordance between results of biopsy and radical prostatectomy specimens.

Gleason Grade Group

More recently there has been a move to change the Gleason Grading system further. It was thought that patients who have Gleason Score 6 disease, the most pathologically insignificant grade of prostate cancer, may not find this nomenclature useful as they may see themselves as being 6 out of 10 on the aggressiveness scale. Pierorazio et al. proposed that different Gleason scores be assigned into different Gleason Grade Groups [30].



This group demonstrated that the grade grouping system accurately predicts prognosis in a group of over 7,000 men who have undergone radical prostatectomy in both the analysis of the Gleason Grade Group of original biopsy results and radical prostatectomy specimens. To date no diagnostic test accuracy study of MRI in prostate cancer has used the Grade Group system.

Pathologically significant and insignificant prostate cancer

The most common used criteria for insignificant prostate cancer in radical prostatectomy specimens is 1) Gleason score of no more than 6 with no tertiary pattern 4 or 5, 2) organ confined disease, 3) tumour volume less than 0.5 cm³ [32]. The use of such criteria has become commonplace as patients with insignificant disease have an extremely low risk of progression to clinical significance in the absence of treatment.

It is important for urologists to have a definition of insignificant disease on biopsy specimens as identification of men with insignificant disease earlier in the diagnostic pathway reduces the risk of overtreatment. The Epstein criteria for insignificant disease on biopsy follows the principles of insignificance on prostatectomy: no Gleason Grade 4 or 5; less than 3 cores from a sextant biopsy containing cancer; no core with more than 50% tumour involvement [33]. Unfortunately, the criteria for insignificance on biopsy does not as confidently predict silent disease as the biopsy does not reliably exclude spread, tumour size and grade. Combining the biopsy results with other parameters such as digital rectal examination (DRE) findings, serum PSA result and other PSA

parameters provides the urologist with more confidence in starting the patient on active surveillance (AS) or surgical or oncological therapy [34].

Prostate cancer risk

The definition of insignificant and significant prostate cancer is a histological criterion and confined by the limitation of the sampling method. The clinical team, lead by urologists, use other parameters to guide treatment. As well as age and comorbidities patients are risk stratified into high, intermediate and low risk based on the stage of tumour (combined DRE and imaging findings), Gleason Score and PSA results as outlined by D'Amico et al. [24] (Table 1-1).

Table 1-1. D'Amico prostate cancer risk groups.

Risk	Parameters
Low	Stage T1-T2a and GS \leq 6 and PSA $<$ 10 ng/ml
Intermediate	Stage T2b and/or GS 7 and/or PSA 10-20ng/ml
High	Stage \geq T2c and/or GS 8-10 and/or PSA $>$ 20ng/ml

Multifocality and index lesion

Prostate cancer is a multifocal disease with more than one cancer focus in over three quarters of radical prostatectomy specimens [35–37]. Wise et al's study of 559 radical prostatectomy specimens found an average of 2.9 separate lesions in each specimen, but 60% of the secondary cancers were less than 0.5cm³ [36]. Not only is prostate cancer multifocal but the foci themselves are typically different grades. Arora et al. demonstrated that the same primary and secondary Gleason grade in each lesion in multifocal disease occurred in only 9% of prostatectomy specimens [35]. This lack of homogeneity between tumour foci grades appears to be due to differences between the foci at a molecular and a genetic level. The TMPRSS2 gene undergoes re-arrangement in the majority of prostate tumours, and Mehra et al. demonstrated there is a different gene arrangement status and class between different tumour foci in multifocal disease, suggesting each focus arises from a different, independent clonal origin [38]. Multiple other

separate gene mutations have been discovered which contribute to the development of prostate cancer and a recent study by Boutros *et al.* found very little shared copy number aberrations and shared nucleotide variants between cancer foci which all points to the existence of multiclonal disease [39–41]. This not only explains the heterogeneity between the aggressiveness of tumour foci within the same gland, but also explains the heterogeneity in the responses of different patients to treatments. This angle of research opens up the potential development of targeted treatments. Diagnosis of prostate cancer may in the future not just be driven by the Gleason Grade, but also by the specific genetic mutations of the tumour so that personalised treatment can be given.

The index lesion is often defined as the tumour nodule, in the setting of multifocal disease, which is the largest in volume [35,36,42], but the latest ISUP guidelines recommend the index lesion should be defined according to the following priority: extra-prostatic extension, then Gleason score, then tumour volume [43]. The largest tumour volume also tends to have the most aggressive staging parameters (such as extracapsular spread, seminal vesicle invasion and neurovascular spread) and also the highest Gleason Score in 89% of cases [44].

Prostate cancer staging

Prostate cancer can be staged clinically, radiologically and pathologically. MRI is the preferred imaging choice for accurate local staging and assessment of pelvic lymph nodes and bony structures. Ultimately, the gold standard for local staging is by histopathological assessment of the radical prostatectomy specimen, but only some patients have a radical prostatectomy and they tend to be younger and have a higher-grade disease. Pelvic lymphadenectomy is sometimes performed at radical prostatectomy providing assessment of the local nodal status. Metastatic lymphadenopathy and visceral metastatic disease can be assessed with computed tomography. Finally, assessment of bone spread with a nuclear medicine technique, known as bone

scintigraphy (bone scan), which is sensitive for the presence of bone metastases. The indications for some of these imaging tests will be covered later.

The American Joint Committee on Cancer (AJCC) cancer staging manual 7th Edition (2009) uses the TNM classification for tumour stage [45]. The 'T' represents the local staging of the 'Tumour', the 'N' represents the 'Nodal' status, and the 'M' represents the assessment for 'Metastasis'. Table 1-2 demonstrates the different TNM components for prostate cancer staging.

Table 1-2. AJCC Prostate cancer TNM staging.

T	Local Tumour Description
TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
T1	Clinically inapparent tumour neither palpable or visible on imaging
T1a	Clinically inapparent tumour in less than 5% of tissue resected
T1b	Clinically inapparent tumour in more than 5% of tissue resected
T1c	Tumour identified on needle biopsy
T2	Tumour confined within the prostate
T2a	Tumour involves one half of one lobe or less
T2b	Tumour involves more than half of one lobe, but not both lobes
T2c	Tumour involves both lobes
T3	Tumour extends beyond the capsule
T3a	Extracapsular extension (uni or bi-lateral)
T3b	Tumour invades seminal vesicle(s)
T4	Tumour is fixed or invades adjacent structures other than seminal vesicles
N	Regional nodal status
NX	Regional nodes were not assessed
N0	No regional nodes involved
N1	Metastasis in regional lymph nodes
M	Distant metastasis
M0	No distant metastasis
M1	Distant metastasis
M1a	Non-regional lymph node(s)
M1b	Bone(s)
M1c	Other site(s) with or without bony metastasis

Biochemical and pathological investigation of prostate cancer

Prostate Specific antigen

Prostate-specific antigen (PSA) is the most commonly used biomarker in cancer diagnostics. A biomarker is defined by the National Institute for Health as a trait that is objectively measured and evaluated as an indicator of biological processes, pathological processes, or pharmaceutical response to a therapeutic intervention [46]. Cancer biomarkers are either produced by the tumour or by the body in response to a tumour. PSA is a serine protease and has been used in the investigation for prostate cancer since the 1980's. Biologically, PSA is responsible for semen liquefaction and is released into the seminal plasma [47]. Its release into the bloodstream is rare in healthy men. Only when there is a breakdown in the basement membrane of the epithelial cells do the serum levels of PSA significantly rise.

Serum PSA can be raised in multiple benign conditions of the prostate, such as benign prostatic hyperplasia (BPH) and prostatitis, following trauma (including from DRE), as well as in prostate cancer. In general, further investigation for prostate cancer is not initiated until the serum PSA level reaches 4mg/ml. However, practice does vary in different institutions as prostate cancer is possible with serum PSA measurements below 4mg/ml and therefore the serum level reflects a continuum with increasing risk of significant prostate cancer with increasing PSA levels [48]. Serum PSA level correlates with both index and total tumour volume [49,50]. Sensitivity and specificity depend on which cut-off is used. Holstrom et al. case-control longitudinal study demonstrated that at PSA levels of 1, 4 and 10ng/ml the sensitivity was 96%, 44% and 15% and the specificity was 44%, 92% and 99% respectively [51]. It is also used in the setting of those with a negative biopsy following initial PSA measurement rises in serum PSA measurement can prompt earlier re-biopsy or imaging of the prostate.

Screening men for prostate cancer using PSA levels has been the source of much debate. The largest randomized screening trial of PSA levels, the European Randomized Study of Screening for Prostate Cancer (ESPRC) using a cut-off of 3ng/ml to prompt prostate biopsy has shown that there is a significant relative risk reduction of 0.80 in the screening arm versus the control for prostate cancer mortality [52]. The disadvantages of population PSA screening are in the significant numbers of false positives and therefore low positive predictive value at PSA levels where tumours are more likely to be lower stage and curable. This leads to many unnecessary biopsies, overdiagnosis of many insignificant prostate cancers, and the morbidity associated with the unnecessary biopsies and treatment. The ESPRC trial estimated that the rate of overdiagnosis was approximately 50% [53]. Recommendations for PSA screening by various institutes and bodies vary around the world. For example the United States Preventative Services Task Force are against screening and the European Association of Urology advise a baseline PSA at 40–45 years [54,55].

Biopsy

Pathological specimens of the prostate can be acquired via three different mechanisms: incidentally at trans-urethral resection of the prostate (TURP), trans-perineal (TP) approach, or via trans-rectal (TRUS) approach which is the most common. Until recently TRUS was most frequently performed in patients with clinical suspicion (DRE and PSA level findings) of prostate cancer then followed by MRI if prostate cancer was detected or the patient had an increased risk of prostate cancer. More recently, there has been a shift in practice to perform MRI before biopsy, which is the order in which other cancers are more commonly diagnosed. This negates the effects of haemorrhage following that biopsy has on interpretation of the MRI. Currently at the Norfolk and Norwich University Hospital (NNUH) TRUS is performed before MRI.

Currently, the most commonly used TRUS approach is the systematic biopsy in which the user takes 6 biopsy specimens blindly from each side of the gland. It is generally untargeted and a

suspicious lesion is not searched for and sampled. Rather, areas of the prostate are focused on and systematically sampled, usually two from the apex, midgland and base. Care is made to not biopsy in the midline to avoid damaging the urethra. The technique is limited to sampling the posterior aspect of the gland missing much of the anterior gland. In general, less than 1% of the gland is sampled at systematic TRUS biopsy [56].

A study of initial and repeat prostate biopsies demonstrated that prostate cancer was found in 34% of initial biopsies and the rate fell thereafter with 25% on second, 24% on third and 21% on 4th or greater [57]. A recent multi-centre paired-cohort study with saturation TP biopsy as a reference standard investigating over 500 men with suspected prostate cancer demonstrated that TRUS biopsy has a sensitivity of 60% for clinically significant prostate cancer indicating that 4 out of 10 significant tumours are missed with TRUS [58]. This in part is due to misses of the biopsy needle of posterior (accessible) tumours, and due to a significant number of tumours (approximately 10%) residing in the anterior gland and likely being inaccessible via TRUS [59].

TRUS prostate biopsy is not without complications. The most common complications are haematuria and pain, but these are usually mild and self-limiting. Infection is possible following the biopsy, however, this is usually limited to a urinary tract infection treatable with antibiotics and occurs in 2-6% of cases [60]. A study of over 75000 prostate biopsies in the UK found that between 2 and 6% of cases were admitted to hospital with complications due to infection [61]. Infection following TRUS biopsy is more common than following TP biopsy, however the incidence of minor and major complication in both groups are similar [62].

TP biopsy can be performed under general or local anaesthetic and is performed by placing a template over the perineum and biopsies taken systematically of the prostate at 5mm intervals. Its advantages over TRUS biopsy is that it can biopsy the entire gland. More recently TRUS and TP biopsy have become targeted by fusing MRI and US images to guide biopsy. The use of image-guided biopsy techniques is beyond the scope of this thesis.

With the shift in practice moving to a biopsy following MRI it is important that MRI is reliably able to identify clinically significant prostate cancers and exclude patients with no or insignificant cancer. Thus, reducing the number of inappropriate TRUS biopsies. Additionally, an accurate and reliable MRI study can also identify identifying patients who would require a TP biopsy rather than TRUS as those with anterior tumours can have a proceed straight to TP biopsy which systematically samples the entire gland.

Treatment of localized prostate cancer

Active surveillance

Approximately 60% of elderly men have prostate cancer on autopsy [63]. The real challenge of treatment of prostate cancer is deciding who and when to treat and to avoid over-treatment, which is an issue in the era of PSA measurement. Active surveillance aims to correctly time treatment rather than provide delayed palliative treatment. The intent is curative treatment, but to minimise treatment-related toxicity without compromising survival and patients should have a life expectancy over 10 years

Active surveillance should be offered to patients within the low-risk group, but they should also have a choice to have radical treatment. There are different criteria used to define low risk, but the European Urology Society suggests that patients who have clinically stage T1 or T2 disease, PSA < 10ng/ml, Gleason score ≤ 6 , ≤ 2 positive cores and a minimal core biopsy involvement of less than 50%. Radical treatment and active surveillance should be discussed with the patient and if it is decided upon they should be followed up with DRE, repeat biopsies and PSA measurement [64].

Radical prostatectomy

Radical prostatectomy is a surgical procedure in which the prostate is removed as well as adjacent fat, the seminal vesicles and the ampulla of the vas deferens. It can be done via a retropubic or perineal approach, as well as open and laparoscopically, and can be robot-assisted. Incontinence (approximately 20%) and erectile dysfunction (approximately 70%) are known complications of the procedure, particularly in the first year [65]. Patients with a life expectancy greater than 10 years can be offered radical prostatectomy with any non-metastatic stage of prostate cancer, but those with locally advanced (T3b/4) or N1 patients should be in highly selective cases. Pelvic lymphadenectomy is an option in intermediate risk cases with a greater than 5% risk of lymph nodes spread and in high risk cases [64].

Focal therapies

High-frequency Ultrasound (HIFU) and cryotherapy are newly available treatments for the treatment of locally staged and small volume prostate cancer. Currently the European Urology Guidelines are that these therapies should not be used outside the trial setting. If these therapies become common place there will be an increased requirement for MRI to aid selection of appropriate patients as smaller tumours suitable for this therapy will need to be visualised and localised within the gland.

External beam radiotherapy

External beam radiotherapy (EBRT) is a form of radical treatment used to treat men with localized prostate cancer and can be used in men with locally advanced disease. It can be used in men who do not have a history of previous pelvic radiotherapy or inflammatory bowel disease. The use of adjuvant hormonal therapies is advocated in intermediate and high risk groups, but the dose of EBRT does not differ dependent on the risk group. Whole pelvic radiotherapy is supported in

those with pelvic nodal disease as well as the long-term use of hormonal therapies, but is not suggested in those with NO disease [64].

Brachytherapy

Low-dose brachytherapy is a viable treatment option for those with a stage of $T \leq 2a$, Gleason Score ≤ 6 with less than 50% cores involved, Gleason score $3+4=7$ with less than 33% of cores involved, a PSA of less than 10ng/ml, a prostate volume of less than 50cm^3 , and an International Prostatic Symptom Score of less than 13 [66]. Radioactive seeds are inserted into the prostate via a trans-perineal approach under ultrasound or MRI guidance. Efficacy of treatment is dependent on adequate coverage of the prostate.

Multiparametric MRI

Prostate MRI protocols consist of T1 and T2 pulse sequences, combined with functional imaging sequences such as DWI, dynamic contrast-enhanced MRI (DCE) and magnetic resonance spectroscopic imaging. The combination of functional and the traditional anatomically-based pulse sequences is known as multiparametric MRI (mpMRI). The current recommendations outlined in the Prostate Imaging – Reporting and Data System version 2 (PI-RADS version 2) [67], the most comprehensive and widely accepted guidelines for prostate MRI, suggest the use of T1, T2, DWI and DCE, but spectroscopy is not part of routine use and there has been a reduction in the volume of research into this technique in prostate cancer. In addition, the use of DCE in prostate cancer is currently controversial and its value in mpMRI was reduced in the PI-RADS version 2 with its use recommended only for the peripheral zone when the DWI findings are indeterminate. Currently at NNUH spectroscopy and DCE are not part of the routine prostate MRI protocol.

T1-weighted imaging

The T1 weighted (T1WI) sequence is a low spatial resolution, large field of view (FOV) sequence performed in the axial plane which covers the entire pelvis from the iliac crests to perineum. The prostate gland on T1WI is homogeneously low intensity with little anatomical detail of the internal architecture of the prostate and no differentiation between peripheral and transitional zones (Figure 1-2). Its strengths are in the assessment of nodal disease in the pelvis, for the presence of bone metastases, and also for haemorrhage within the gland which manifests itself as hyperintensity on T1 weighted images.

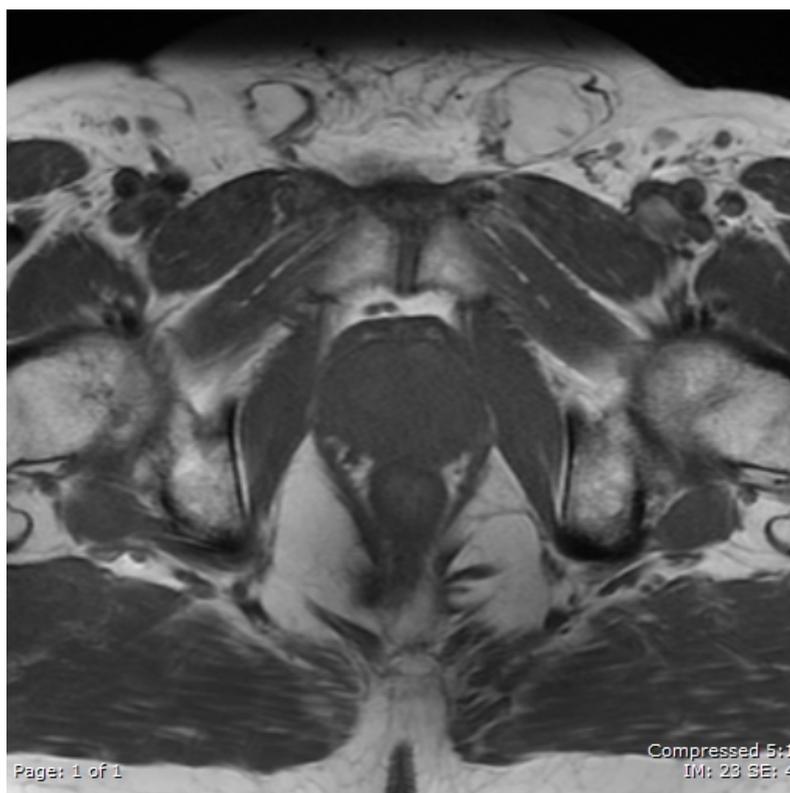


Figure 1-2. T1WI of the prostate demonstrating the homogenous low signal of the gland and lack of differentiation of PZ and TZ.

Haemorrhage occurs within the prostate following a biopsy of the gland. This can result in an area of hypointensity on T2-weighted imaging (T2WI) thus mimicking tumour, increasing the risk of false positives, and limiting the interpretation of the MRI. Following biopsy haemorrhage can spread throughout the ductal system and involve a greater volume of prostate than expected based on the location of the biopsy cores. As most biopsies are performed via a trans-rectal approach the majority of the haemorrhage is located in the peripheral zone and seminal vesicles.

The duration of delay from biopsy to MRI is controversial. The current recommendation is for a delay of 6 weeks [67,68], however, this is often considered too long when cancer treatment targets need to be met and can add to patient anxiety. The volume of haemorrhage within the gland determined subjectively on T1WI is inversely proportional to time from biopsy [69–71], however, it is its effects on accuracy which are often the reasons cited for reducing the delay to MRI following biopsy. Rosenkrantz et al. (2012) qualitative assessment of accuracy of multiple sequences at detecting prostate cancer found a trend towards reduced sensitivity of T2WI, but

DWI and DCE accuracy parameters were maintained and furthermore, there was no significant effect on accuracy of delays following biopsy [72]. A further study by Rosenkrantz et al. (2010) assessing the quantitative parameters of regions of haemorrhage on tumours and normal tissue, found that although the ADC value of tumour increased if affected by haemorrhage there was still a significant difference in ADC value of tumour compared to normal tissue [73].

Currently at the NNUH the compromise between pressure to meet cancer treatment targets and allowing the haemorrhage to reduce following biopsy is the imposition of a 28 day delay from the day of biopsy to MRI. In keeping with PI-RADS version 2 guidelines T1WI of the pelvis are used to assess for post-biopsy haemorrhage [67].

T2-weighted imaging

Protons within water molecules within the body align when they are in a magnetic field in a process known as longitudinal magnetisation. When a radiofrequency pulse is applied to the protons they fall out of alignment in a process known as transverse magnetisation. When the pulse ends the transverse magnetisation decreases and disappears. This occurs because the protons interact with adjacent protons, exchange energy and result in a loss of phase coherence across a population of protons in a process known as T2 relaxation. The interaction of the protons and exchange of energy is different for protons within different tissues.

The T2 value for different tissues is the time for transverse magnetisation to decay approximately 37% of its initial value. Protons within free water such as in fluid or glandular tissue has less interaction and exchange of energy and therefore has much slower relaxation time, but water within less water-based tissues (tumour, muscle, fibrous tissue) have more interactions and shorter T2 values.

The time to echo (TE) is the time interval between the 90° radiofrequency pulse and the point where signal is recorded from the imaged field. Given tissues have different T2 values and signal

obtained from different tissues will vary depending on the TE, the TE can therefore be changed to optimise the signal intensities and thus contrast between tissues. Often with a greater TE there is a greater contrast between tissues but this comes at the expense of continually dropping signal to noise ratio (SNR) from the entire image making it potentially non-diagnostic.

The T2 sequences used come into two categories; spin and gradient echo sequences, and the former can be conventional or fast spin echo sequences. The fast spin echo sequence is typically used in prostate imaging with the advantage of this technique being its speed of acquisition, but can come at the expense of reduced signal. As mentioned the TE is chosen to optimise contrast between tissues and in the prostate it is important to delineate between peripheral and transition zone as well as between normal prostate tissue and tumour. For MRI of the prostate The T2 values of prostate cancer, transition zone and peripheral zone have been calculated at 1.5T as 82, 88 and 122ms [74] with the TE of the fast spin echo sequence set usually between 100 and 130ms to maximise tissue contrast.

The T2WI sequence is the imaging mainstay of any MR prostate protocol. Current recommendations are for high resolution T2WI images in three orthogonal planes (axial, coronal and sagittal) with 3mm slices and no gap between slices [67]. Currently, the three-plane high resolution T2WI sequences in the NNUH MRI prostate protocol are performed at 3.6mm with no gap.

The advantage of the T2WI sequence is the high level of anatomical detail of the prostate and surrounding structures combined with the ability to distinguish prostate cancer from normal prostate tissue.

McNeal's anatomical divisions of the prostate are well demonstrated in the axial plane (Figure 1-3). The peripheral zone is demonstrated as a homogenous hyperintense crescent of tissue extending from base to apex. The anterior fibromuscular stroma is a hypointense band which

extends down the anterior aspect of the prostate and is in continuity with the detrusor muscle of the bladder. The transitional zone, which is a small proportion of the gland in younger men, becomes an increasing fraction of the gland with advancing years due to the development of benign prostatic hyperplasia (BPH). BPH arises in the transitional zone and can extrude into the peripheral zone. It is usually seen as hyperintense cystic nodules, but stromal nodules are low signal and both nodule types can exhibit mixed signal intensities. The central zone resides at the base of the gland and is seen as a dumbbell-shaped area of low signal housing the ejaculatory ducts. The ejaculatory ducts extend caudally from the base of the seminal vesicles. The seminal vesicle are paired lobular areas of hyperintensity on T2WI extending cranial from the base of the gland. The prostatic capsule is separated from adjacent structures by a thin low signal rim and if tumour is seen to extend beyond, this then changes the staging from T2 to T3 disease.

Pathology within the prostate appears as low signal in comparison to normal glandular tissue. In the peripheral zone, tumours are appreciated as a round or ill-defined area of low signal, but has a significant amount of overlap with other conditions and states such as scars, prostatitis, haemorrhage, and atrophy [75]. The detection of cancer within the transitional zone is a greater challenge as there is a greater overlap of the appearances of tumour and the transitional zone, particularly in the presence of BPH. Lesions are typically homogeneously hypointense, uncircumscribed, with a lentiform or spiculate shape [67].

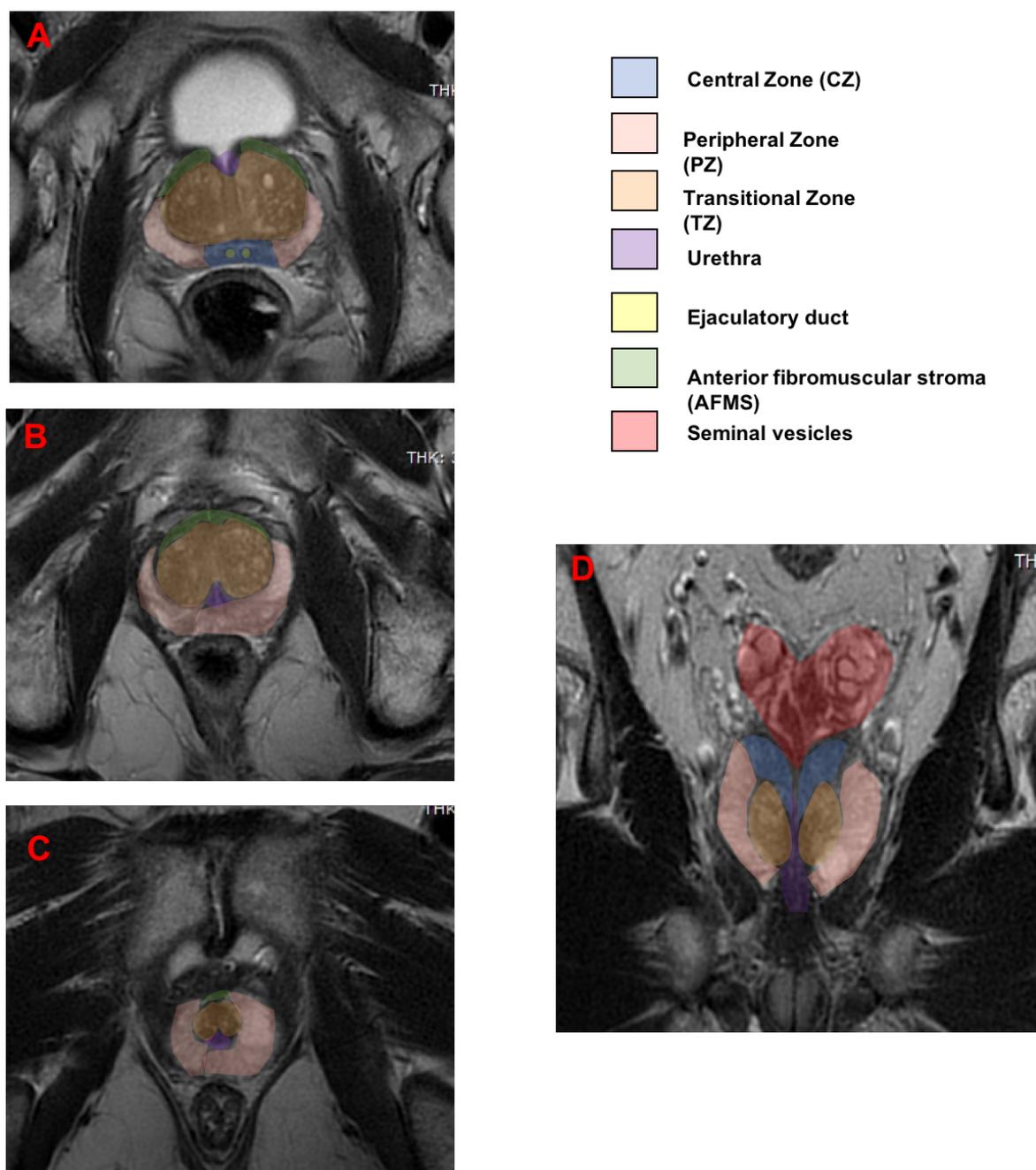


Figure 1-3. T2WI of a normal prostate demonstrating the anatomical divisions of the prostate and seminal vesicles. A - Axial slice through the base of the gland. B - Axial through the mid-gland. C - Axial through the apex. D - Coronal slice.

Diffusion-weighted imaging

Conventional MRI is based on the ^1H signal from water ($^1\text{H}_2\text{O}$) which has the advantage of large volumes throughout the body providing sufficient signal to form an image. Contrast in conventional imaging is based on how the physical properties of water ^1H behave in different tissues and chemical environments [76]. For example, in the brain the T2 relaxation time for

cerebrospinal fluid is large providing an abundance of signal, which is in comparison to brain parenchyma which has a shorter relaxation time and displaying less signal thus providing an inherent contrast between the two tissues.

Diffusion-weighted imaging (DWI) also uses the ^1H from water to provide signal, but instead of using how water behaves in different environments it works on the principle of water movement.

In the body water molecules move randomly in a pattern known as Brownian motion. DWI studies the movement and displacement of water during the gap between two diffusion-sensitising gradients. Water movement between the gradients leads to attenuation of signal between the gradients. Therefore, in tissues where movement of water is 'free', such as within fluid filled structures like the urinary bladder, there is a greater loss of signal. In comparison, tissues with increased cellularity and intact cell membranes create greater impedance to water movement and thus restricted diffusion (Figure 1-4). Tumour tissue in general has higher cellularity and greater intact cell membranes [77] and therefore greater restricted diffusion.

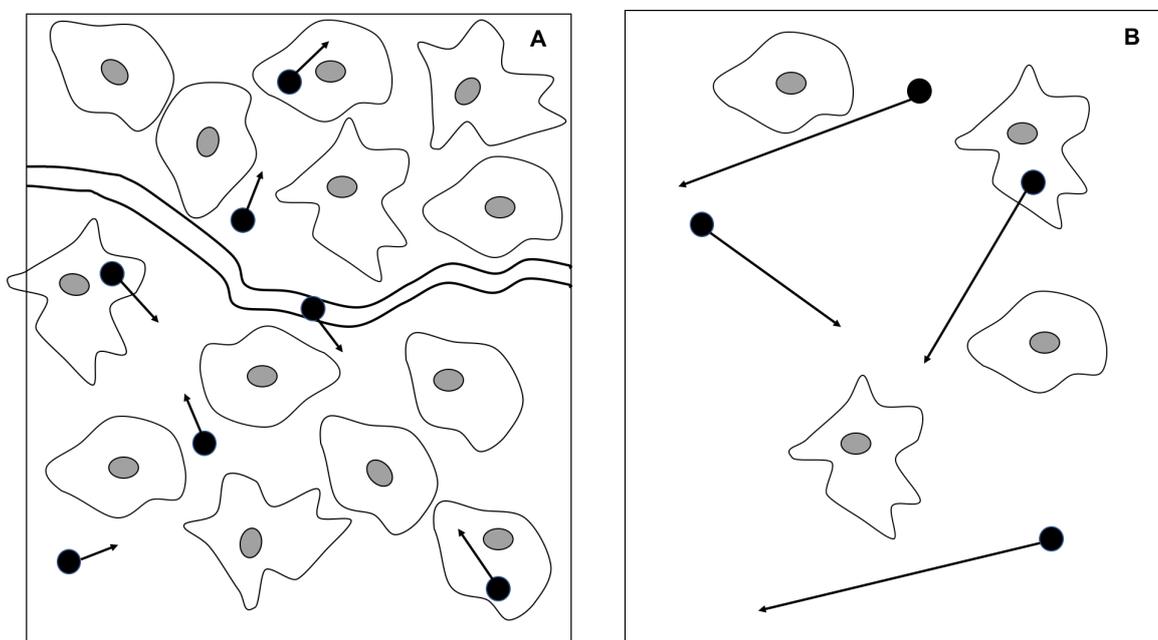


Figure 1-4. Diffusion of water molecules. A - Highly cellular environment in which the diffusion of water is restricted due to impedance by the extracellular space and cell membranes. B - Free movement of water in large extracellular space or due to faulty cell membranes.

It is important to note that the MRI signal in DWI is sensitive to the diffusion of water in the extracellular and intracellular space, and also the intravascular space. Water molecules within the intravascular space tend to move further than those in the intra and extra-cellular spaces and the contribution of the intravascular water movement to DWI signal varies between tissues. This variation also occurs in tumours, in which the intravascular water diffusion of highly vascular tumours can contribute significantly to the DWI signal [78].

The principles of DWI were initially described by Stejskal and Tanner [79] who designed an experiment that detected and quantified the diffusion of water molecules in vivo. It was used on a standard T2WI sequence in which an identical pair of diffusion-sensitising gradients were applied to the tissue around a 180° radiofrequency refocusing pulse. Figure 1-5 demonstrates the difference in diffusion of static and moving water molecules during the image acquisition and how they result in different signal output. Static molecules acquire phase information from the initial diffusion-sensitising gradient and when the information is rephased during second diffusion-sensitising gradient there will be minimal reduction in the measured signal intensity. Conversely when the second diffusion-sensitising gradient is applied to tissues with mobile water molecules there is no reversal of the phase shift and less rephasing of water molecules leading to a reduced signal. Therefore, the attenuation of signal is proportional to the degree of water movement.

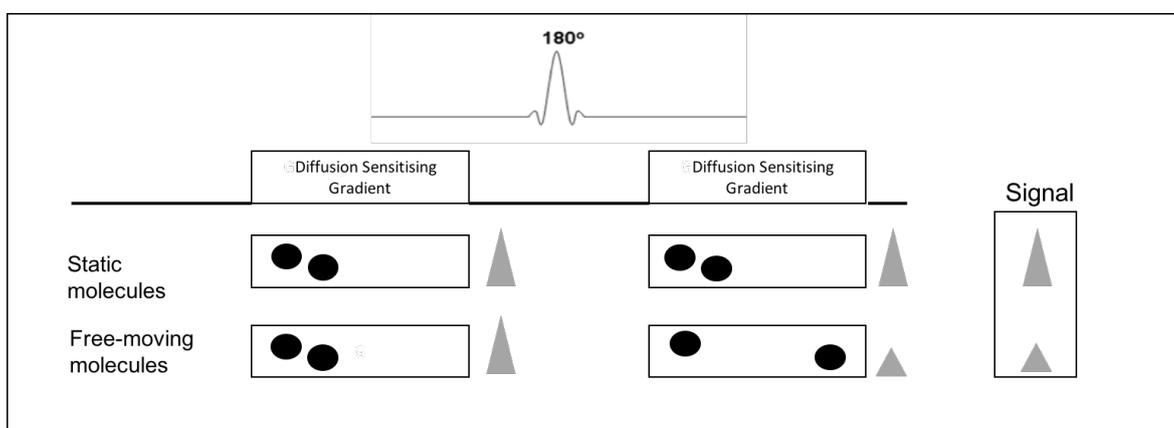


Figure 1-5. Water diffusion metrics demonstrating the attenuation signal in different tissues between the diffusion sensitising gradients with static molecules, such as in tumour, maintaining signal and free-moving molecules, such as the bladder, losing signal.

The sensitivity of the water molecules to the DWI sequence depends on the amplitude of the diffusion-sensitising gradient, the time between the gradients and the duration of the diffusion-sensitising gradient. The 'b-value' is a measure of the magnitude of the diffusion weighting provided by the diffusion gradients, usually the gradient amplitude rather than the duration of and interval between gradients, and is expressed in sec/mm^2 . Diffusion gradients can be applied to any type of sequence, however, currently they are applied to echo-planar imaging (EPI) sequences as they are fast and solve many of the problems of motion artefacts.

The apparent-diffusion coefficient (ADC) is the measure of the diffusion. It is calculated mathematically from at least two b-value images, often more. The signal attenuation, a marker of diffusion, of each voxel is plotted against b-value with the slope of the line representing the ADC and measured in mm^2/sec . Each unit of ADC equates to a signal intensity and therefore, the ADC values can be used to create an ADC map, a set of images which is a visual representation of the 'true' diffusion of the scanned area in which each voxel holds information on the amount of restricted diffusion within. On the ADC map an area of restricted diffusion will be demonstrated as low signal and correspond to high signal on the DWI images. By drawing a region of interest on the ADC map the ADC value of the included area can be calculated, thus providing a quantifiable measurement of diffusion.

In clinical practice, when running DWI sequences on an area of the body the DWI images, including all b-values, are present to visually and qualitatively assess, as is the ADC map. With the ADC map also providing means of quantifiably measuring diffusion.

Many parameters of the DWI sequences can be altered to optimise the images. An issue with DWI is that it is prone to low signal. This is in part is due to signal decay due to T2 effect in the sequence. Signal continues to fall as the b-value is increased because the sequence time increases as the parameters of the diffusion-weighted gradients change to increase the b-value. Therefore, other mechanisms need to be in place to increase the gradient strength whilst reducing the

acquisition time. EPI sequences are used as they generally lower the acquisition time whilst maintaining SNR.

DWI in prostate MRI

Of the three functional sequences in mpMRI of the prostate, DWI is used most frequently and relied most heavily on in current reporting guidelines [67]. Its appeal in prostate cancer detection is the increased cellular density of tumour compared to normal prostate tissue leading to an inherent contrast between tumour and non-tumour tissue, the lack of need for exogenous contrast agents, and relatively quick acquisition makes it a routine part of the mpMRI protocol. In addition, it has potential uses as a non-invasive biomarker of tumour aggressiveness [80].

One of major technical factors to be researched within DWI of the prostate is the choice of b-values to use. The advantages of using higher b values are an increased tumour to non-tumour contrast made achievable by greater suppression of signal of normal prostate tissue and theoretically improved tumour detection. However, this is traded against reduced signal to noise ratio (SNR), and increased susceptibility and motion artefacts [81]. Therefore, at higher field strengths the benefits of greater SNR permit possible use of higher b values.

The issue with lower b-values is that the displayed signal intensity echoes the effects of tissue diffusion and the T2 weighting in the image rather than just the amount of restricted diffusion. Normal prostate tissue can exhibit high signal on T2WI reflecting its high glandular components, thus lacking contrast with tumour which will be high signal due to the restricted diffusion. This obscuration of tumours can be seen even with b-values up to 1000 sec/mm² with a study demonstrating that over half of tumours were not visible at b1000^a sec/mm² using a 1.5T MRI despite the reader being provided with a histopathology tumour map to aid localisation [82].

^a It is common practice to refer to the b-value used as b1000 sec/mm², for example, rather than stating 'a b-value of 1000 sec/mm².'

Rosenkrantz et al. [81] separately assessed b-value images and ADC maps created using both b1000 and b2000 sec/mm² images at 3T with radical prostatectomy correlation and demonstrated that using a b-value of 2000 sec/mm² compared with a b-value of 1000 sec/mm² resulted in significantly improved tumour sensitivity and higher tumour-to-peripheral zone (PZ) contrast on the DWI images, but these benefits were not observed in ADC maps created with maximum b-values of 2000 sec/mm² compared to 1000 sec/mm². Ueno et al. [83] also reported significant benefits in specificity, accuracy, and positive and negative predictive values with b-values of 2000 sec/mm² compared to b1000 sec/mm² for detection of prostate cancer in the PZ, but was unable to demonstrate these findings for tumours of the transitional zone (TZ). Tumour detection has also been shown to be better at b2000 sec/mm² by other studies at 3T [84,85], and also at 1.5T [86].

A few studies have compared more than b1000 and b2000 sec/mm² sequences. Metens et al. [87] reported that tumour contrast to noise ratio and image quality were improved in b1500 sec/mm² images compared to b-values of 1000 and 2000 sec/mm² in both central gland and PZ tumours. These results were also supported by Wang et al. [88] who found significantly higher SNRs at b1500 than 1000 and 2000 sec/mm². Contrast between normal tissue and tumour was best at b1500 and significantly greater at b2000 and b1500 than with b1000 sec/mm². When these b-values were used to create individual ADC maps and the ADC values of lesions were used to distinguish the likelihood of tumour by assessing the area under the curve (AUC) of the receiver operating characteristic (ROC) curve the ADC map created from b1500 data was significantly higher, indicating greater accuracy, than a ADC map from b2000 for both PZ and TZ, and ADC map from b1000 sec/mm² for the TZ. However, the ADC map from the b1000 data outperformed the map from b2000 sec/mm² in both zones of the prostate.

Wetter et al. [89] also analysed multiple b values and concluded that the signal intensity of tumour and normal tissue decline with increasing b value but the decline is more pronounced in

normal tissue. The contrast ratio at $b1500$ and $b2000 \text{ sec/mm}^2$ were significantly better than at $b1000$ and $b800 \text{ sec/mm}^2$ but there was no significant difference between $b1500$ and $b2000 \text{ sec/mm}^2$.

When ADC maps, created using different b-values, were assessed in isolation, Wetter et al. [89] concluded that the contrast ratios of tumour using ADC maps from multiple b-values (800 , 1000 , 1500 , and 2000 sec/mm^2) were not significantly different. The reduced usefulness of the ADC maps derived from ultra high b values is suggested by Tamada et al. [84] with further studies demonstrating improved diagnostic performance of ADC maps derived from $b1000$ compared to $b2000 \text{ sec/mm}^2$ [90,91]. Conversely, using quantitative analysis, ADC maps obtained with a b-value of 2000 sec/mm^2 were superior to that obtained with a b-value of 1000 sec/mm^2 for PZ tumours but less so for TZ cancers [92].

Traditionally, the lower b-value used to create the ADC map was 0 sec/mm^2 . This has been replaced by a low b-value between 50 and 200 sec/mm^2 as this reduces the effects of the early capillary component on the measured diffusion signal. This was based on work by Le Bihan et al. [93] in which two compartments (tissue and capillary) rather than one in each tissue are recognised. In this model, there is pseudo-diffusion in the capillary compartment due to the rapid movement of water molecules. The effects of this are reduced by using a higher minimum b-value and then fitting the mono-exponential decay curve to calculate the ADC value of tissue. Thörmer et al. [94] performed a study assessing the effects of using different b-values to create different ADC maps, examined the ADC values of tumour and normal tissue, and compared lesion conspicuity of different ADC maps. When ADC maps of $0,800 \text{ sec/mm}^2$ were compared to those using $50,800 \text{ sec/mm}^2$ the ADC values of tumour did not change significantly, but lesion conspicuity was improved with the latter. The ADC maps on the MRI protocols at NNUH are all creating with a non-zero b-value, specifically 100 sec/mm^2 .

The diagnostic accuracy of DWI and the methodology of diagnostic test accuracy studies of prostate MRI are examined in the systematic reviews and meta-analyses in Chapter 2 (Page 55) and Chapter 3 (Page 82).

Multiple studies have proven the ADC value of tumour is significantly lower than that of benign tissue [80,90,95–98]. However, amongst these studies the ADC values of tumours are variable, as is the degree of overlap with normal tissue. For example, the mean ADC values of tumour in Kim et al. (2010) [90] was $820 \times 10^{-6} \text{ mm}^2/\text{sec}$ compared to Kim et al. (2007) [99] who found the mean tumour ADC to be $1,300 \times 10^{-6} \text{ mm}^2/\text{sec}$. The variation is dependent on multiple factors, most notably the choice of b-values used. Several studies have shown that the ADC value of tumours decreases with increasing b-values [83,89,94]. Given the large discrepancy in ADC measurements of tumours the current recommendations for the assessment of tumour with ADC maps is in the visual assessment of tumour with DWI images and ADC map rather than by measuring ADC values [67].

The evidence suggests that b-values of over $1000 \text{ sec}/\text{mm}^2$ should be used to detect tumour, particularly at 3T, but when it comes to creating ADC maps the use of b-values of 1000 outperform ADC maps created with a b-value of $2000 \text{ sec}/\text{mm}^2$, however, the use of ADC maps with $b1500 \text{ sec}/\text{mm}^2$ data may be of benefit. This is reflected in the PI-RADS version 2 recommendations which advocate using a high b-value of $1400 \text{ sec}/\text{mm}^2$ or more, but the highest b-value used to create the map should be $1000 \text{ sec}/\text{mm}^2$ [67]. The issue with much of the higher b-value studies thus far has been the lack of diagnostic accuracy studies rather than the use of, for example, lesion conspicuity, contrast ratios and SNR of lesions, as used in many of the studies above. Furthermore, the meta-analyses of diagnostic test accuracy of DWI alone [100] and DWI and T2WI [101] have not attempted to extract the b-values chosen in the included studies and compare if possible the pooled findings of different groups. Performing this extraction and

analysis in a meta-analysis would provide stronger evidence as to the practical and clinical benefits of using higher b-values.

With increasing Gleason Grade there is an increase in cellular density within the tumour with a resulting decrease in the free movement of water and more restricted diffusion. This results in a decrease in ADC value with increasing Gleason grade, a correlation which has been described in multiple studies [102–106]. The correlation coefficients of mean ADC to Gleason grade for these studies range from weak ($\rho = -0.31$) to moderate ($\rho = -0.66$) with a considerable overlap of ADC values with different Gleason grades.

The b-values chosen to make the ADC map has an influence on the ability to distinguish different grades of prostate cancer. Peng's [107] study of 11 different ADC maps created with using different combinations of b-values found the lowest correlation ($\rho = -0.30$) with an ADC derived from b1500,2000 sec/mm² and a highest correlation ($\rho = -0.68$) from b0,1500 sec/mm². However, there were wide confidence intervals in this study and overlap in correlation coefficients between b-values typically used in routine clinical work. Other studies directly comparing ADC maps from very high b-values with lower b-values have found no significant difference between them.

Kitajima et al. [92] reported that the correlation of ADC value and tumour aggressiveness (Gleason score) was fair ($\rho = -0.33$) and Tamada et al. [84] demonstrated no significant difference in detecting high or intermediate risk lesion with b0,2000 sec/mm² vs b0,1000 sec/mm² ADC maps, however, there was a trend towards that conclusion.

The studies described above have described and focused on mean tumour ADC value, however, others have used different ADC metrics other than the mean to determine whether there is a stronger correlation. Wu et al. [108] found a stronger correlation using the minimum ADC value rather than the mean. Donati et al. [103] establish the 10th percentile ADC value to have the strongest correlation ($\rho = -0.36$) with the mean and median ADC value having coefficients of -0.31 and -0.30 respectively.

Tumour ADC values have potential to help discriminate tumours of different grades as more aggressive tumours have a lower ADC value than lower grade tumours. However, like the use of tumour ADC for discriminating tumour from non-tumour tissue the use of ADC values clinically to determine the grade of disease is not used in routine clinical practice. The underlying principles behind the correlation are sound and it is important when investigating novel DWI techniques that this correlation is assessed and compared against conventional DWI and the literature.

Novel DWI techniques

As well as increasing the highest b-value other techniques have been created to improve the accuracy and reliability of DWI imaging of the prostate. Most of these techniques have revolved around improving the spatial resolution of the scan. Medved et al. [109] found by reducing the slice thickness and improving the in-plane resolution there was a significant improvement in qualitative measures of image quality, lesion conspicuity, and internal architecture definition. In addition, they found the ADC values of tumours with the higher resolution DWI sequence to be higher than for the standard DWI sequence, however, the ADC maps for each sequence were calculated using a different number and choice of b-values, which is known to have a significant effect on ADC value [107].

Another technique recently employed is to reduce the field-of-view of the image. Traditionally, DWI imaging of the prostate has been limited to the imaging of the whole pelvis, in which the prostate occupies only a small volume of the whole imaged pelvis. Within a larger field of view there are magnetic field inhomogeneities, and distortions due to susceptibility artefact, which are increased at 3T, all of which decrease image quality. The small cross-sectional size of the prostate within the larger field of view limits the spatial resolution. By reducing the field of view the aim is to reduce the off-resonance induced artefacts by reducing the read-out duration.

This is achieved by using a 2D spatially selective radiofrequency excitation pulse, followed by a 180° refocusing pulse which selects only a selected area for imaging. The result is a reduction in the echo-train-length which minimises distortion caused by susceptibility artefact and also reduced the number of k-space lines in the phase encoding direction which enables higher resolution [110,111].

The prostate is a challenging organ to image with DWI as the bladder anteriorly and the gas containing rectum posteriorly means the imaging is prone to susceptibility artefact and geometric distortion. Reducing the field of view theoretically has the potential to improve both.

Furthermore, larger slice thickness and field-of-view also have reduced spatial resolution which can result in the missing of smaller tumours. Ultimately, the aim of small field-of-view DWI is to improve the detection and localization of prostate cancer, with the improved spatial resolution also permitting better tumour delineation and therefore more accurate staging. With any new DWI sequence the ADC value of tumours and normal prostate tissue must be determined and the correlation of tumour ADC and tumour grade performed to assess the potential for the sequence to non-invasively assess tumour aggressiveness.

Reducing the field-of-view has been demonstrated to be advantageous in the clinical setting in the spinal cord [112], pancreas [113] and breast [114]. Within the prostate variations of reduced field-of-view DWI imaging has been used by multiple groups [110,115–118]. There is consistency amongst the results of the reduced field-of-view studies of the prostate in that there is improvement in the qualitative measures of image quality and distortion, with some groups demonstrating a reduction in quantitative measures of distortion [110,119]. When ADC was measured in normal and tumour tissue there is variation in ADC values found between the large and reduced FOV techniques with some finding that values are higher with reduced FOV [116,117], some lower [118] and some found no significant difference between techniques [119]. Brendle et al. [119] and Feng et al. [115] are the only groups to assess for prostate cancer, with

the former the only to test for diagnostic accuracy, and both groups use biopsy specimens as a reference standard for all or some of the included patients. Brendle et al. [119] did not demonstrate a significant difference in sensitivity, specificity or AUC between the reduced and large FOV technique, but it is unclear if they tested for multifocality and only used 15 patients.

Field-of-view optimized and constrained undistorted single-shot (FOCUS) is a commercially available small FOV DWI sequence created by General Electric. Feng et al. [115] used the sequence in their assessment of prostate cancer concluding the small FOV technique to have improved image quality, and reduced distortion, blurring and artefact with higher ADC values of tumour when compared to the conventional large FOV technique. They used b-values of 0 and 800 sec/mm² to qualitatively assess the image quality and to create the ADC maps. The limitations were that they only assessed the index lesion, used biopsy specimens as the reference standard and had a small number of patients.

Molecular imaging of prostate cancer

Molecular imaging of prostate cancer is an emerging field that aims to provide further non-invasive information about tumour biology, behaviour and diversity not just in the prostate, but across the whole body. There are a variety of methods of assessing prostate cancer at a molecular level and they assess cell metabolism, membrane proteins, and hormone receptors. Most techniques have concentrated on the combination of one these techniques with anatomical imaging such as CT. Most notably the combination of radiolabelled choline, a nutrient essential for the synthesis of cell membranes during cell growth, and radiolabelled prostate specific membrane antigen, a membrane glycoprotein found at 100-1000 times higher proportions in prostate cancer, with CT [120]. These are examples of molecular imaging of cell metabolism and membrane receptors, respectively, and choline PET is available in some centres within the NHS.

The limitation of using the above techniques is the poor spatial resolution locally within the prostate, so the detection of smaller lesions and local staging is poor. MRI spectroscopy is a technique which measures the levels of metabolites, which are altered in prostate cancer, thus becoming a biomarker for malignancy. In normal prostate citrate levels are high and choline low, whilst in prostate cancer the reverse is observed. Creatine is another metabolite measured, and although the levels between tumour and normal tissue are similar, the choline and creatine-to-citrate ratios are normally observed. MRI spectroscopy is technically challenging, performed in few centres, and achieving the imaging acquisition and interpretation quickly enough to be used routinely is not possible. For these reasons spectroscopy is not routinely used in prostate MRI.

MRI can also be fused with PET and can be used locally to detect and stage prostate cancer. Gallium labelled prostate-specific membrane antigen PET/MRI has been investigated and a recent study has shown it to have improved accuracy at prostate cancer detection compared to PET alone and also mpMRI [121]. The combination of molecular imaging with MRI is promising, but is currently a long way from being available routinely with its use limited to trials.

Diagnostic test accuracy studies

To optimise clinical practice, it is important for radiologists and other members of the clinical team to understand the strengths and weaknesses of the tests they request, report and base clinical decisions on. Awareness of the accuracy and reliability of these diagnostic tests is vital for safe management of patients. Diagnostic test accuracy studies are performed to assess the ability of a test to correctly identify a patient with a disease and also exclude disease from patients without a condition, however, they are prone to particular biases, and the studies with particular shortcomings can lead to the overestimation of accuracy of tests [122,123].

The Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement was published in 2003 to provide guidance on the reporting of diagnostic accuracy studies to improve transparency

and comprehensiveness [124]. Since its creation there has been an increase in the reporting of the 25 items which were deemed the minimum set of essential items to be included in a publication of a diagnostic test accuracy study. Although, there are still deficiencies in certain aspects, in particular the reporting of inclusion criteria, blinding and confidence intervals [125]. There is, however, higher reporting of the items in accuracy studies of radiological tests in comparison to laboratory tests [125].

In 2015, there was an update to the STARD reporting item list, resulting in 30 essential items to report in diagnostic test accuracy studies [126]. The new list now includes, for example, the requirement to describe the study hypotheses to keep the focus of the results narrow and avoid generous interpretation of the results. In addition, prospective studies should be considered as trials and registered on a clinical trials registry to allow identification of the study's existence and prevent selective reporting.

It is the intention with the FODIP^b study to recognise and report the 30 items, in particular, the reporting of biases and limitations with the aim of answering the clinical questions outlined in the methods and materials (Page 109) and avoid making over-reaching generalisations when discussing the results and comparing to the current literature.

Choice of reference standard and correlation with MRI

Diagnostic test accuracy studies of MRI in the detection of prostate cancer are varied and attempt to answer many clinical questions. The focus of the accuracy studies can be on the detection and localisation of prostate cancer within the gland, or the presence and extent of extracapsular extension or seminal vesicle invasion.

^b FODIP refers to the name of diagnostic test accuracy study of this thesis.

The choice of reference standard and patient group has a significant impact on the type of study. Using biopsy samples as a reference standard allows investigation of a pre-treatment population including men potentially with no prostate cancer. However, the limitations of using TRUS biopsy specimens is the significant number of false negative studies due to the small volume of tissue in the biopsy sample compared to total prostate volume and the non-targeted approach [56].

Ahmed's [58] multicentre paired validating confirmatory study assessing the diagnostic accuracy of mpMRI versus TRUS biopsy in patients with a clinical suspicion of possible prostate cancer and using systematic high volume transperineal core biopsy specimens of the entire gland found the sensitivity of TRUS to be 48% for clinically significant prostate cancer.

The use of transperineal biopsies as in the above study could be considered the best way to sample the prostate in a clinical and research environment as the entire gland is sampled, clinically significant tumours are less likely to be missed and tumours can be localised to certain aspects of the gland. However, the limitations are an inability to accurately measure tumour dimensions, the procedure usually involves general anaesthetic, requires considerable resource, and has significant morbidity, particularly given a reasonable proportion of patients will not have any or at most clinically insignificant prostate cancer.

The use of radical prostatectomy specimens as the reference standard has benefits and weaknesses. The radical prostatectomy specimen can be sampled in a variety of ways, but the most common and recommended method, is by whole mount section of the entire gland which provides axial slices through the prostate, except the extreme apex, which are close in alignment to MRI [127]. This potentially allows for the closer correlation of MRI and prostatectomy specimens; including lesion localisation within the gland using landmarks both visible in pathology specimen and on MRI such as urethra [95]; lesion size and shape; delineation and increased confidence in the presence of multiple tumours; and assessment of spread beyond the capsule or into the seminal vesicles. The use of radical prostatectomy specimens does introduce selection

bias to diagnostic test accuracy studies as the patients who have this surgery tend to be younger and fitter, and their disease is more likely to be of higher grade and therefore more likely to be visible at MRI. In addition, the correlation is not exact as slice thicknesses and locations may not match and they may not both be in the same plane. Finally, after assessment of MRI and radical prostatectomy, preferably independent of each other and blinded to all clinical details, the correlation must be performed by a person(s) which introduces human error into the determination of the accuracy of the MRI sequence.

STARD's most recent iteration [126] introduced definitions of "Key STARD terminology". The guidelines differentiate between 'Clinical Reference Standard', which is the best available method for establishing the presence and absence of the target condition, and 'Gold-standard', which is an error free reference standard. The use of radical prostatectomy specimens as the reference standard introduces many biases and limitations, and although the alternatives allow the assessment of a population with and without prostate cancer, the ability to correlate better with axial MRI slices and consistently assess for all tumours makes radical prostatectomy specimens the best clinical reference standard for diagnostic test accuracy studies of MRI sequences.

The PROMIS study and comparison to FODIP

One of the most influential pieces of research affecting the management of patients with suspected prostate cancer was the PROMIS study [58]. As mentioned, this was a prospective multicentre paired validating confirmatory study assessing men with suspected prostate cancer. In this study 576 men with suspected prostate cancer underwent a prostate MRI before a TRUS biopsy. This was followed by a non-targeted systematic transperineal biopsy, which was used the reference standard. The important aspect of the prostate MRI was that it was unaffected by haemorrhage (as it was done by before biopsies); was on a 1.5T magnet (to make the study relatable as all institutions will have a 1.5T magnet, and not necessarily a 3T magnet); the protocol used DCE in addition to DWI and T2WI; the MRI readers received training prior to participating;

and then had clinical details (such as PSA and DRE) available. They assessed diagnostic accuracy of the MRI and TRUS biopsy against the TP biopsy as reference standard, using different thresholds for a positive sample based on percentage of the biopsy core involved and Gleason grade of tumour.

The results showed that MRI was significantly more sensitive, but less specific than TRUS biopsy for all definitions of prostate cancer, however, most importantly the negative predictive value of MRI was significantly greater than TRUS. The advantage of a high negative predictive value for MRI is that if you can confidently exclude patients with significant prostate cancer you can prevent the man going on for an unnecessary TRUS biopsy, whilst still having the high sensitivity required to diagnose a significant number of prostate cancers. The results of the study are significant in that it is the most robust piece of evidence that there is a benefit in switching the investigative pathway of men with suspected prostate cancer from TRUS biopsy then MRI if positive to MRI then TRUS or TP biopsy.

There are very few similarities between the FODIP and PROMIS studies. FODIP is a diagnostic accuracy study of a specific DWI sequence on a 3T magnet with blinded readers, comparing their reads against blinded reads of radical prostatectomy specimens by histopathologists. It is therefore difficult to directly compare the results of both studies, however, the significance of the PROMIS study's results and the shift to pre-biopsy MRI does prompt a change in how the results of the FODIP study are considered, particularly how the DWI sequences will fit into a pre-biopsy setting as this is how almost all investigate suspected prostate cancer now.

Chapter 2 Meta-analysis: Accuracy of high b-value DWI

Background

The standards outlined in the PRISMA statement [128] provide a framework for conducting a systematic review and meta-analysis. The reporting standards and checklist are not exclusive to reviews of diagnostic accuracy, but are also relevant to cost-effectiveness, interventions and therapeutics, policy making and prognostic questions. Parameters described in the checklist include systematic review registration and publication of review protocol, eligibility criteria, study search and selection, data collection, risk of bias, presentation and synthesis of results, limitations and conclusions.

Systematic reviews and meta-analyses of diagnostic test accuracy are important for clinicians. They offer an evidence-based display and summary of a potentially large body of evidence whilst minimizing bias and improving the reliability of conclusions [129]. They can provide summaries of sensitivity, specificity, positive and negative predictive values, odds ratios and summary receiver operating characteristic curves [130]. Crucially, they can help decide whether results are inconsistent or can be generalized.

The diagnostic accuracy of mpMRI and DWI for the detection of prostate cancer is well-established, with multiple meta-analyses reporting their diagnostic performance [100,101,131]. Current guidelines for the use of DWI in prostate MRI emphasise the use of b-values greater than 1000 sec/mm² [67], but none of the published meta-analyses have attempted to extract information on the ideal b-value.

Accuracy of DWI in prostate cancer can be performed in different ways, namely a qualitative, visual assessment for tumour using either or both DWI and ADC map images, or using a quantitative, ROI based method of assessment for tumour using cut-offs of ADC values.

Visual assessment of the diagnostic performance of prostate MRI can be achieved through different methodologies. The most common approaches are a lesion-based and a sector-based accuracy assessment.

Lesion-based analysis

The lesion-based analysis is perhaps considered the more clinically appropriate where readers identify lesions within the prostate and are asked to localise them to a part of the prostate. Readers are either required to identify just the index lesion or can assess for multiple lesions. The definition of the index lesion often varies, but is usually the largest tumour and highest Gleason score on pathological specimens, usually prostatectomy [132]. Readers can be assessed 'stringently' or 'approximately'. A 'stringent' approach requires the reader to correctly identify the tumour in the exact location within the prostate that it is identified on histopathology. For the 'approximate' approach, if the radiologist correctly identifies the lesion near the exact location it is considered a true positive assessment. Both approaches have merit. The 'stringent' accuracy may be of increased importance when the results are used for targeted biopsy or focal therapies, however, the 'approximate' approach reflects the limitations of correlating histopathology and radiology.

Turkbey et al. [133] described a 'stringent' and 'approximate' approach, with the latter considering positive radiologic-pathologic correlation if the tumour was localised to a neighbouring sector. Rosenkrantz et al. [134] modified Turkbey's [133] 'approximate' lesion localisation method by suggesting positive radiologic-pathologic correlation was present when a lesion in the neighbouring zone was chosen as long as the correct side of the gland (right or left) and zone (peripheral or transitional) was selected.

The lesion-based analysis has been used to test the accuracy of multiple different imaging techniques in prostate cancer detection [81,134–137], however, the limitation is that areas

without prostate cancer are not assessed thus preventing the 'True Negatives' being identified and preventing the completion of the 2x2 contingency table. This limits the analysis of the diagnostic performance of the technique to sensitivity and positive predictive value and preventing the assessment of specificity and negative predictive value. Therefore, precluding the use of these studies in a diagnostic test accuracy meta-analysis as a complete 2 x 2 contingency table is an eligibility requirement.

Sector-based analysis

The sectoral based analysis allows 'True Negatives' to be assessed and the completion of the 2 x 2 contingency table. This facilitates other diagnostic performance parameters, such as area under the receiver operating characteristic curve and accuracy, to be calculated.

For this methodology, the prostate is divided into a specific number of sectors. Radiologists then decide if each sector has or does not have tumour present or provides a probability, usually a Likert score, of likelihood of tumour presence in each sector. The pathologists then assess the prostatectomy or biopsy specimen and map the tumour out on a similar reporting template to the radiologist. The different cut-offs available when using a Likert scale allows the interplay between sensitivity and specificity to be gleaned thus allowing the ROC curve to be created and AUC calculated. Determining the absence of tumour is an important finding when reading prostate MRI. A drawback to this method is that the analysis is at the mercy of the limitations in the radiologic-pathologic correlation and thus follows the 'stringent' criteria outlined in the lesion-based analysis above. This could falsely lower the calculated diagnostic performance of the sequence.

There is a variation in the literature as to how many sectors the prostate is divided into. Most biopsy papers limit the number of sectors to between 6 and 8, but when the reference standard is a prostatectomy specimen the prostate is divided into 8 sectors [138–141] or 12 sectors

[104,142]. In both approaches the peripheral zone is divided into left and right in the gland base, mid-gland and apex. The difference is in the division of the transitional zone, in which the 8 sector approach divides the whole transitional zone into left and right, but the 12 sector approach considers the 3 different levels like in the peripheral zone. The 12 sector approach equalises the importance of identifying tumour in all three levels of the prostate as well as in the peripheral or transitional zone.

A balance must be struck between dividing the prostate into enough regions to allow a fair assessment of tumour localisation, but have too many and the assessment is falsely low as the challenge of identifying the sector becomes greater than the identifying the tumour.

High b-value meta-analysis

The aim of this systematic review and meta-analysis was to determine the diagnostic performance of high b-value DWI at detecting prostate cancer. In addition, it was important to recognise whether there were any differences in the qualitative and quantitative assessment of DWI accuracy. A subgroup analysis of different study characteristics could highlight whether any parameters of study design, MRI protocolling or method of assessing for tumour affect the accuracy of high b-value DWI.

The meta-analysis also provided an up-to-date measure of the accuracy of high b-value DWI and allowed comparison with the diagnostic performance of the DWI sequences used in the FODIP study as well as permitting a comparison with other published meta-analyses. Additionally, it had the potential to provide information on common sources of bias and limitations in similar accuracy studies which possibly could then be minimised within the FODIP study methodology.

Material and Methods

This meta-analysis was reported using the preferred reporting items for systematic reviews and meta-analyses outlined in the PRISMA statement [128]. The review was registered prior to commencing on PROSPERO, an international database of prospectively registered systematic reviews in health care. The PROSPERO reference number is CRD42015027644 and the protocol is found at http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015027644. The status of the meta-analysis was kept up-to-date throughout.

Search strategy

A comprehensive systematic literature search was independently performed by two reviewers (KG, TSy) to identify studies investigating the diagnostic accuracy or performance of high b-value DWI for detecting prostate cancer. A MEDLINE search is presented in Table 2-1. In addition, searches were conducted of EMBASE, and the grey literature/trial registry databases: WHO International Clinical Trials Registry Platform and OpenGrey. Studies were not limited by country of origin, but were limited to those published in English. All searches were from database inception to 1st January 2016.

Table 2-1. MEDLINE search strategy

	Search	Number of studies
1	"prostatic neoplasms/diagnosis"[MeSH Terms]	54849
2	"diffusion magnetic resonance imaging" [MeSH Terms]	13053
3	1 AND 2	402
4	Limit 3 to Humans	398
5	Limit 4 to English	372
6	Limit 5 to Abstracts	351

Eligibility criteria

Retrospective and prospective studies were included if they reported detection of prostate cancer in a pre-treatment population using high b-value DWI of the prostate. Only primary research articles, available as full-text, were accepted; however, review articles were checked for additional primary references. High b-value was defined as a $b_{\max} \geq 1000 \text{ sec/mm}^2$. Histopathological results as a reference standard (biopsy or radical prostatectomy), and sufficient data to calculate true positive (TP), false positive (FP), false negative (FN), and true negative (TN) data were required. A minimum of 10 patients was also required. If multiple b-values were used, including $b < 1000 \text{ sec/mm}^2$, the study was only eligible if data with $b \geq 1000 \text{ sec/mm}^2$ could be extracted. Studies using high b-value DWI in combination with other diagnostic sequences to detect cancer were excluded.

Study identification

Titles and abstracts from the search results, and the full-text papers for all studies which met or potentially met the eligibility criteria, were independently reviewed by two reviewers (Keith Godley (KG), Tom Syer (TSy)). Each reviewer performed the study search independently with the pre-determined eligibility criteria as guidance. An abstract or title which potentially met eligibility criteria resulted in the assessment of the full text. At this point both reviewers discussed their eligible papers. Those studies which met the eligibility criteria on full review were included in the final analysis. Disagreements on inclusion suitability were resolved by consensus between the two reviewers (KG, TSy).

Data extraction

Both reviewers then independently extracted the data on a pre-defined template, including: publication year, country of origin, sample size, description of study population (age), study design (prospective, retrospective, or unknown), patient enrolment (consecutive or not), inclusion

and exclusion criteria, reasons for exclusions from analysis, and number of experts who assessed and interpreted MRI results. Data were recorded on: blinding of MRI measurements to clinical, biochemical or histopathological results; methods used to determine diagnosis; types of coils; and b-values used. For each study, the number of true-positive, false-positive, true-negative, and false-negative findings for high b-value DWI in diagnosing prostate cancer was recorded. Disagreements in data extraction findings were resolved through discussion or through adjudication with a third reviewer (Toby Smith (TSm)). In some instances, not all the data for the 2 x 2 contingency table was published, however, if the missing data could be calculated from other data, for example, positive predictive value or specificity results, then the study was included. No attempts were made to contact authors to ask for missing data.

Quality Assessment

Two reviewers (KG, TSy) independently assessed each included paper's quality using QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) [122]. QUADAS-2 is a publically available resource designed to provide a standardised quality assessment of different parameters of diagnostic test accuracy studies and a modified version has been endorsed by the Cochrane Collaboration and advised for use in any diagnostic test accuracy Cochrane reviews. The results of each reviewer's QUADAS assessment were discussed by the two reviewers and disagreements were resolved through discussion or through adjudication with a third reviewer (TSm).

Statistical analysis

Study heterogeneity was assessed through examination of the data extract table. On review of the principle characteristics of the included studies, such as patient population, study methodology, imaging protocol, and reference standard used were broadly homogenous, meaning a meta-analysis was appropriate.

Specificity and sensitivity of each study was calculated using 2×2 contingency tables. Pooled sensitivity, specificity and positive and negative likelihood ratios with 95% CIs were calculated. Finally, the specificity and sensitivity were used to calculate a summary receiver operating characteristic (sROC) curve and the area under the curve (AUC). This was performed using the Moses-Littenberg method which has been used extensively in diagnostic test accuracy meta-analyses [143].

Statistical heterogeneity between the included studies was assessed using the Cochran's Q test, Q , and the inconsistency, I^2 [144]. When $p = < 0.10$ and $I^2 = > 50\%$, unexplained statistical heterogeneity was evident and diagnostic performance analyses were performed using a random-effects model. These parameters are performed using the below software and they are of increased benefit in the assessment of statistical heterogeneity in therapeutic and intervention meta-analyses, but as they are designed to assess heterogeneity around a single outcome variable they are less benefit in diagnostic test meta-analyses as they do not allow for the influence of threshold effect [145]. Statistical heterogeneity is almost always expected in diagnostic test accuracy meta-analyses and random-effects model of assessment of diagnostic performance is performed as a default.

Threshold effect is an important concept when reviewing the results of a diagnostic test accuracy meta-analysis. Different studies assessing for the same condition can use a different threshold/cut-off for a positive test. The accuracy of the study therefore depends on the choice of threshold and changing the threshold can increase sensitivity and reduce specificity, or vice versa. Differences in threshold used between studies can result in differences in their summary data, a concept known as threshold effect.

Threshold effect can be assessed visually and statistically. To do so visually this needs to be done with both the sensitivity and specificity data of all studies present simultaneously. This is performed using a Receiver Operating Characteristic (ROC) curve where sensitivity is plotted

against (1-specificity) and all the studies are included. If threshold effect is observed the studies will all be grouped in a 'shoulder-arm' shape. Quantitatively assessing the correlation between sensitivity and (1-specificity) using a Spearman correlation coefficient of the logit of sensitivity and the logit of (1-specificity) allows statistical assessment of threshold effect, with a strong correlation ($p = <0.05$) indicating the heterogeneity between studies could be explained by threshold effect [145].

Threshold effect was not found to be factor in the heterogeneity between studies and therefore a meta-regression and subgroup analysis were performed to explore other sources of heterogeneity to determine how they influence diagnostic performance. Meta-regression of pre-determined co-variates, such as reference standard and threshold method, was computed using a generalisation of the Littenberg and Moses model allowing weighting for random effects and by the inverse of variance or sample size [143,146].

All statistical computations were performed using Meta-DiSc (version 1.4, Javier Zamora) and Review Manager (version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). The statistical analysis and creation of tables and figures was performed by KG.

Results

Search results

A summary of the search strategy results is presented in Figure 2.1. In total, 351 studies were identified from the search results, of which 61 were deemed potentially eligible. After full-text review, 10 studies met the final eligibility criteria and were included in the analysis [90,91,141,142,147–152].

Characteristics of included studies

The principle characteristics of the included studies are displayed in Table 2-2, with imaging and study methods listed in Table 2-3. From the 10 included studies, 522 patients were analysed, with mean (range) age = 64 (43–87) years. The mean and median PSA was 19 and 9.3 ng/mL, respectively.

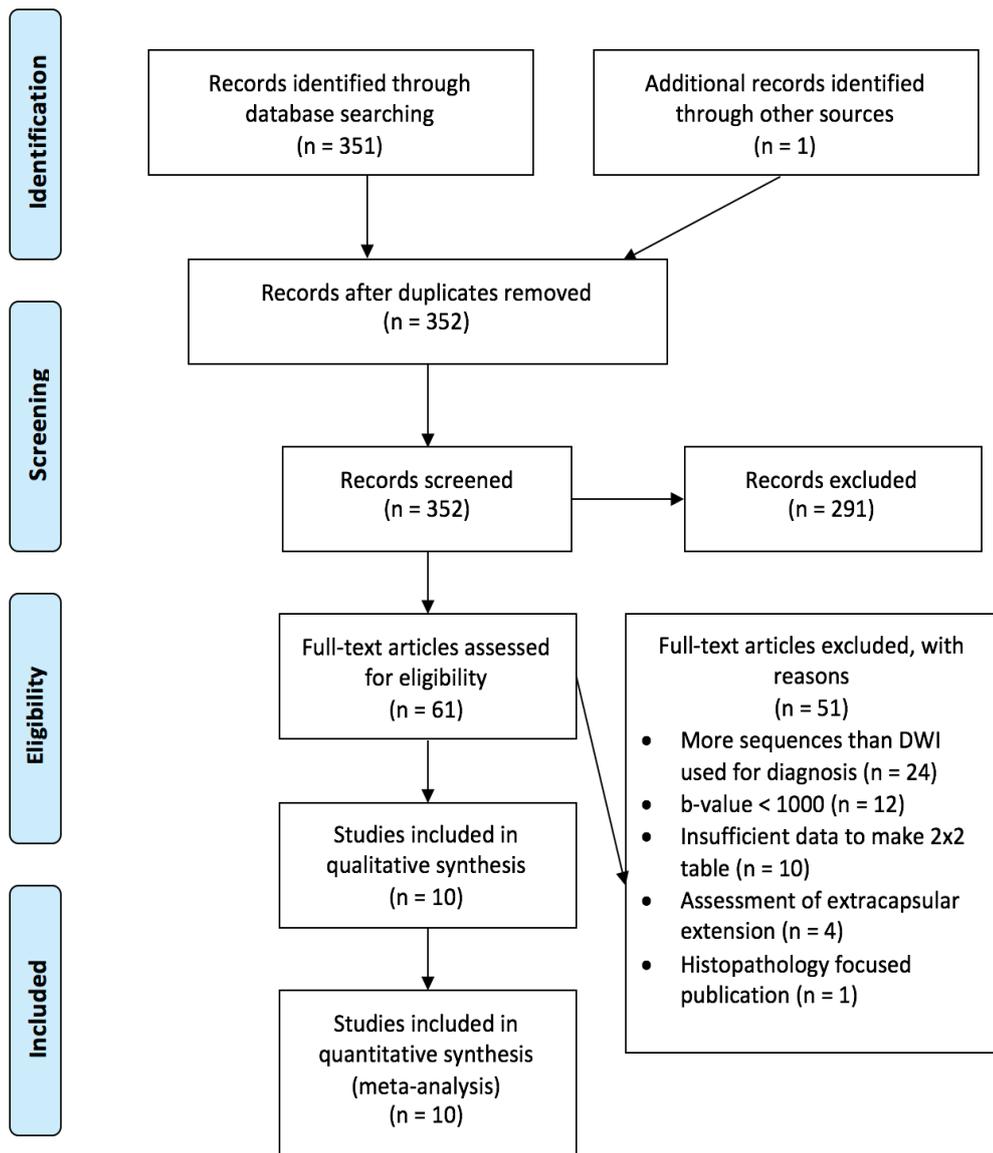


Figure 2-1. Flow diagram of study identification and exclusion.

Table 2-2. Principle characteristics of included studies.

Study	Year	Country	Number. of patients	Mean Age (range)	Mean PSA (range)	Design
Chen [147]	2008	China	42	63 (45-82)	52.5 (4.7-147)	Retro
Girometti [148]	2012	Italy	26	64*(51-74)	6.0*(2.5-10)	Pro
Isebaert [141]	2013	Belgium	75	66*(49-74)	10.4 (1.5-70.9)	Pro
Kim [90]	2000	South Korea	48	66 (45-80)	7.2*(2.3-23.2)	Retro
Koo [91]	2013	South Korea	80	66 (45-81)	7.2 (1.2-57)	Retro
Kumar [149]	2007	India	23	64.5	11*(0.5-1000)	Pro
Lim [142]	2009	South Korea	52	65 (48-76)	66 (45-80)	Retro
Peng [152]	2013	USA	48	61.5 (44-73)	15.6 (0.8-256)	Retro
Rosenkrantz [150]	2015	USA	58	63	8.2	Retro
Vilanova [151]	2011	Spain	70	63.5 (43-87)	7.4*(4-17.2)	Retro

PSA – Prostate specific antigen (ng/mL), * Median, Pro – prospective, Retro – retrospective

Table 2-3. Imaging and methodological characteristics of the included studies.

Study	Field Strength	Coil	b-values (sec/mm ²)	TE (ms)	Spasm	Reference Standard	Blinded	Threshold method
Chen	1.5T	A	0 & 1000	94	U	Bx (sextant)	Y	ADC
Girometti	3T	A	0, 800 & 1200	60	Y	Bx (octant)	U	ADC
Isebaert	1.5T	A	0, 50, 100, 500, 750 & 1000	79	U	RP (octant)	Y	Visual
Kim	3T	A	0 & 1000; 0 & 2000	83-95	Y	RP (14 sectors)	Y	Visual
Koo	3T	A	0 & 300; 0 & 700; 0 & 1000; 0 & 2000	75-76	Y	RP (10 sectors)	Y	Visual
Kumar	1.5T	B	0, 250, 500, 750 & 1000	96	U	Bx (whole gland)	Y	ADC
Lim	3T	A	0 & 1000	117	Y	Bx (12 sectors)	Y	Visual
Peng	1.5T	B	0, 50, 200, 1500 & 2000; 0 & 1000*	71-85	Y	RP	N	ADC
Rosenkrantz	3T	A	50 & 1000	86	U	RP (4 sectors)	Y	ADC
Vilanova	1.5T	B	0 & 1000	94	U	C (octant)	Y	ADC

Coil A – without endorectal coil, *Coil B* – with endorectal coil; *TE* – Echo time; *RP* – radical prostatectomy, *Bx* – biopsy, *C* – both RP and Bx included; Y – yes, U – unclear; *Spasm* – Anti-spasmodics; *ADC* – Apparent diffusion coefficient

* 29 patients were imaged with b-values of 0,50, 200, 1500 and 2000; 24 patients were imaged with b-values of 0 and 1000.

Three studies were prospective and seven retrospective. Field strengths of 1.5T [141,147,149,151,152] and 3T [90,91,142,148,150] were each used in five studies. Radical prostatectomy specimens were used as the reference standard in six studies [90,91,141,142,150,152], biopsy specimens in three [147–149] and one study used a combination [151]. The MRI reader was blinded in eight studies [90,91,141,142,147,149–151], blinding was not known in one, and one was not blinded [152]. Anti-spasmodic agents, either glucagon or hyoscine butylbromide, were used in five studies [90,91,142,148,152] and their use unknown in the remainder. Five studies [90,91,142,148,152] used b-values of greater than 1000 sec/mm².

Several methods were used to detect prostate cancer: region-of-interest (ROI) based ADC quantification was used in six studies [147–152]; and visual assessment of lesions was performed in four studies [90,91,141,142]. A more detailed description of the ADC quantification and visual assessment threshold methodology is found in Table 2-4 and Table 2-5, respectively.

All ten studies used the monoexponential function to create ADC maps and estimate ADC. In 3 studies [90,91,150], extracted data were split into subsets. Kim et al. [90] and Koo et al. [91] generated multiple ADC maps: b = (0, 1000), and (0, 2000) sec/mm² for the former, and b = (0, 300), (0, 700), (0, 1000), and (0, 2000) s/mm² for the latter. Rosenkrantz et al. [150] split data into peripheral and transitional zone results. The other studies generated only one set of ADC maps from their DWI data, performing monoexponential fitting to all acquired b-values

Table 2-4. Description of threshold method used for included studies using ADC measurements

Study	Description of threshold method used
Chen	ROI measurement of visually low signal area and 'normal' tissue creating a 5-point scale from the quintiles of the ADC measurements.
Girometti	ADC cut-off value of $0.9 \times 10^{-3} \text{ mm}^2/\text{sec}$.
Kumar	Mean ADC for all the peripheral zone and central gland. Gland assessed as a whole.
Peng	10th percentile ADC measurements of non-blinded assessment of the ADC map.
Rosenkrantz	Normalised ADC measurement of lowest signal area in PZ and TZ of both lobes.
Vilanova	ROI measurement of octants with 5-point scale using ADC cut-offs.

Table 2-5. Description of threshold method used for included studies using visual assessment of tumour.

Study	Description of threshold method used
Isebaert	Visual assessment of DWI and ADC map. Binary cut-off.
Kim	Visual assessment of ADC maps. Binary cut-off.
Koo	Visual assessment of DWI and ADC. Likert scale.
Lim	Visual assessment of ADC map. Likert scale.

Quality assessment

Study quality assessment is presented in Table 2-6. Figure 2-2 demonstrates the QUADAS-2 graphical summary of the studies indicating the proportion of high, low, or unclear risk in each domain. A high risk of bias was demonstrated in the patient selection domain, but overall the quality of the studies included was considered ‘good’.

Table 2-6. Quality assessment of the included studies.

	Risk of bias				Applicability concerns		
	Patient Selection	Index test(s)	Reference Test	Flow & Timing	Patient Selection	Index test(s)	Reference Test
Chen	+	+	-	+	+	+	+
Girometti	-	+	-	-	-	+	+
Isebaert	-	+	+	U	+	+	+
Kim	-	+	+	+	-	+	+
Koo	-	+	+	+	-	+	+
Kumar	+	+	-	+	+	+	+
Lim	-	+	+	+	+	+	+
Peng	-	-	+	U	-	+	+
Rosenkrantz	-	+	+	U	+	+	+
Vilanova	+	+	-	+	+	+	+

+ Low risk, – High risk, U Unclear risk

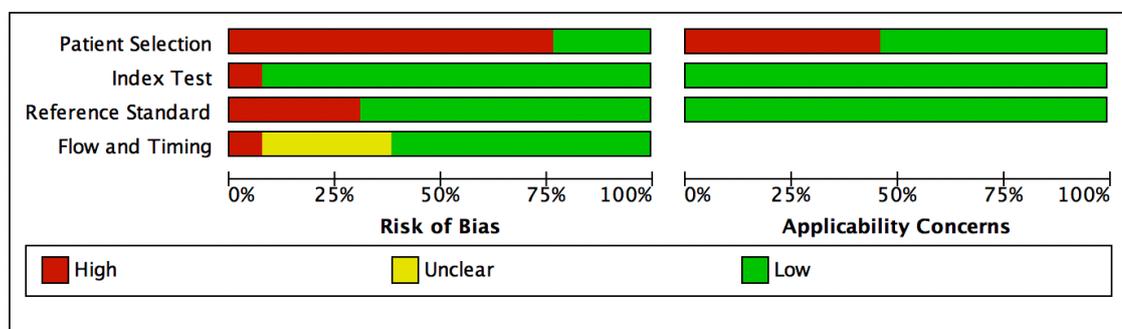


Figure 2-2. QUADAS-2 results summarising the proportion of low, high and unclear risk of bias and applicability concerns

Diagnostic performance

The results from the individual studies are presented in Table 2-7. The pooled sensitivity and specificity of high b-value DWI MRI in detecting prostate cancer was 0.59 (95% CI, 0.57–0.61; Figure 2-3) and 0.92 (95% CI, 0.91–0.92; Figure 2-4), respectively. Sensitivity and specificity heterogeneity tests gave $Q = 435.05$ ($p \ll 0.001$), $I^2 = 97.2\%$ and $Q = 89$ ($p \ll 0.001$), $I^2 = 86.5\%$ respectively, indicating significant statistical heterogeneity between studies.

Table 2-7. Diagnostic performance of eligible studies and subsets.

Study	TP	FP	FN	TN	Sens	Spec	Notes
Chen	42	37	9	164	0.82	0.82	
Girometti	4	14	8	182	0.33	0.93	
Isebaert	359	44	617	732	0.37	0.94	
Kim	158	49	22	443	0.88	0.92	b=1000
	128	40	52	452	0.71	0.92	b=2000
Koo	174	38	31	557	0.85	0.94	b=1000
	152	22	53	573	0.74	0.96	b=2000
Kumar	17	10	6	27	0.63	0.82	
Lim	171	57	56	340	0.75	0.86	
Peng	49	6	12	37	0.86	0.80	
Rosenkrantz	15	13	24	64	0.39	0.83	TZ
	42	2	42	30	0.50	0.94	PZ
Vilanova	37	8	14	81	0.73	0.91	

TP – true positives, FP – false positives, FN – false negatives, TN – true negatives, Sens – sensitivity, Spec – specificity, TZ – transitional zone, PZ – peripheral zone

The pooled positive and negative likelihood ratios for high b-value DWI MRI in detecting prostate cancer were 6.64 (95% CI, 4.9–9.0; Figure 2-5) and 0.33 (95% CI, 0.2–0.5; Figure 2-6), respectively. Positive and negative likelihood ratio heterogeneity tests gave $Q = 82.50$ ($p \ll 0.001$), $I^2 = 85.5\%$ and $Q = 517.45$ ($p \ll 0.001$), $I^2 = 97.7\%$, respectively, indicating significant statistical heterogeneity between studies.

Figure 2-7 shows the sROC curve of the 10 studies, where AUC = 0.92, indicating ‘good’ diagnostic accuracy [153].

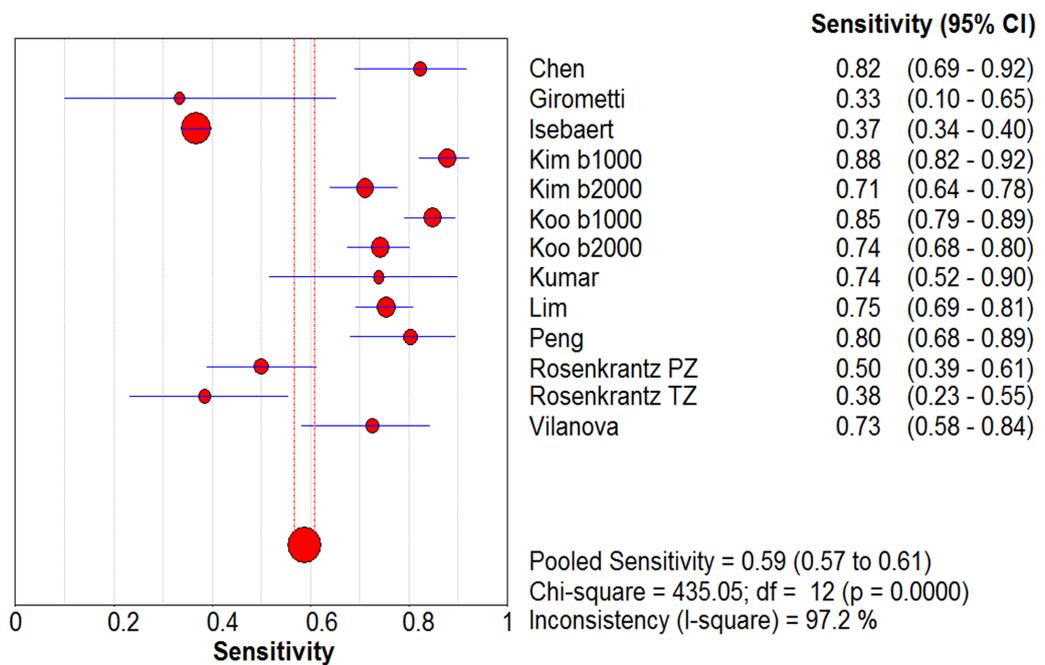


Figure 2-3. Forest plot of sensitivity with pooled sensitivity, Q statistic of the chi-squared, and I-squared results.

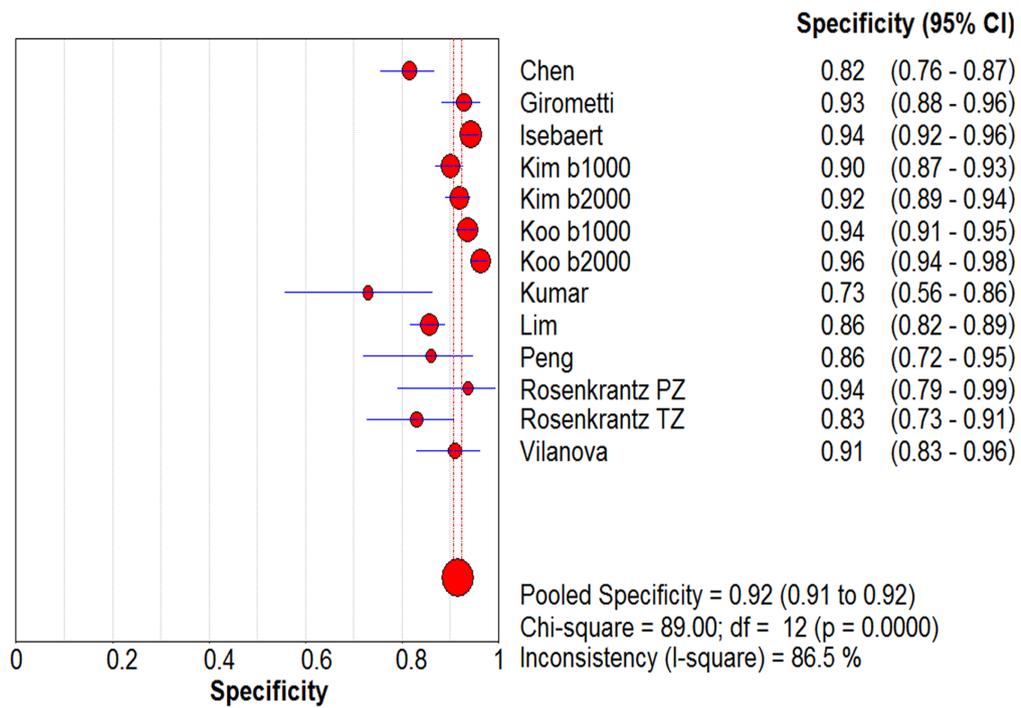


Figure 2-4. Forest plot of specificity with pooled specificity, Q statistic of the chi-squared, and I-squared results.

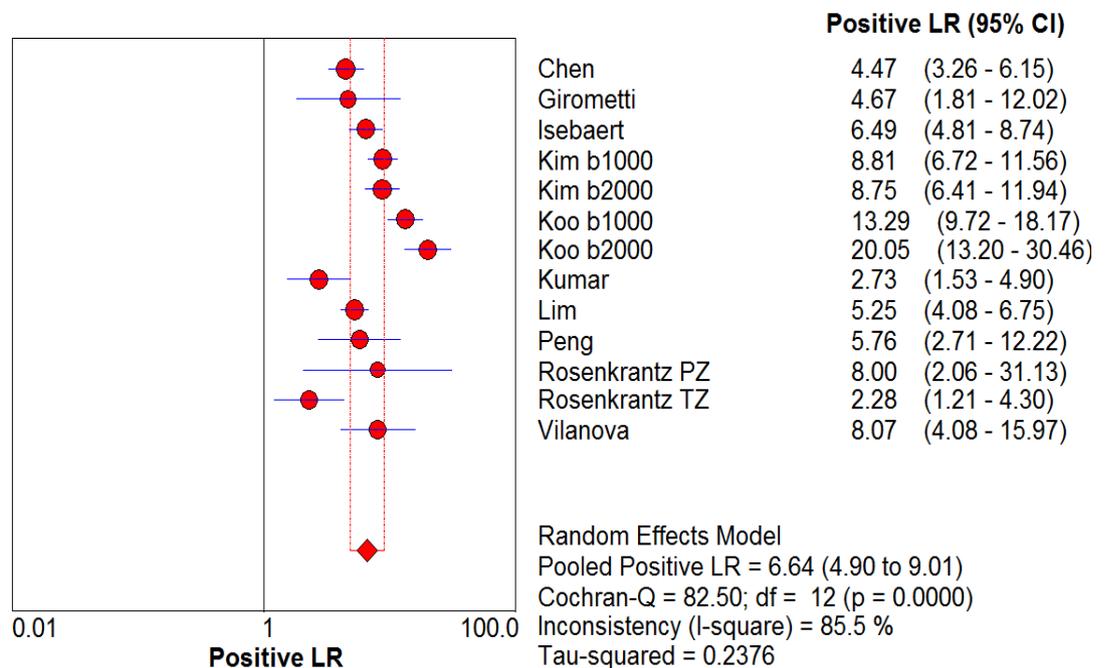


Figure 2-5. Forest plot of positive likelihood ratio with pooled positive likelihood ratio, Q statistic of the chi-squared, and I-squared results.

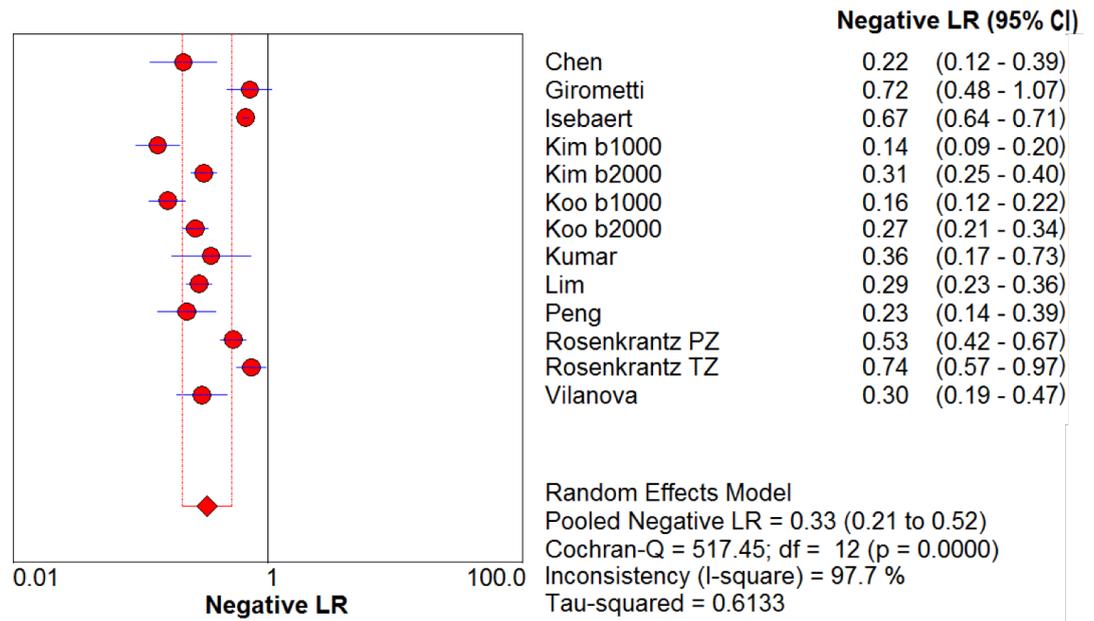


Figure 2-6. Forest plot of negative likelihood ratio with pooled negative likelihood ratio, Q statistic of the chi-squared, and I-squared results.

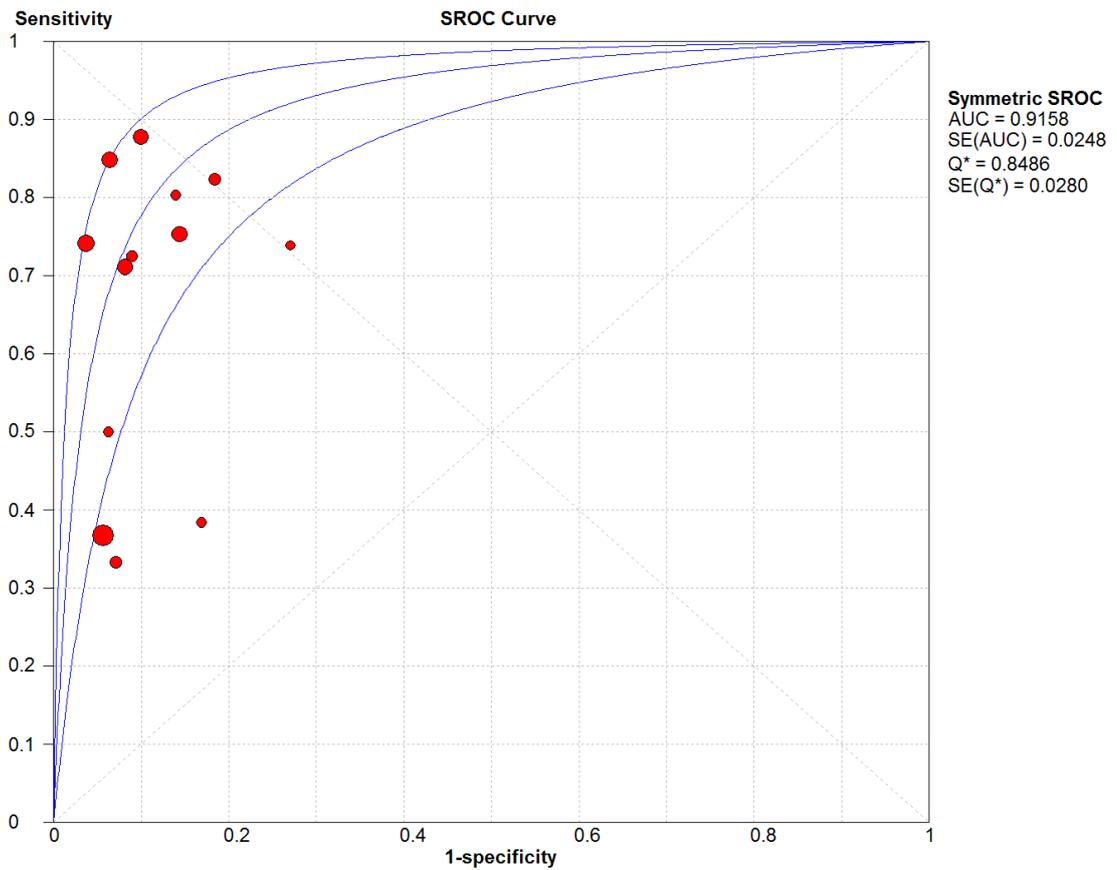


Figure 2-7. The summary Receiver Operating Characteristic (sROC) curve for high b-value DWI in detecting prostate cancer.

Meta-regression analysis

The ROC curve did not demonstrate a ‘shoulder-arm’ shape (Figure 2-8) and the Spearman Correlation Coefficient between the logit of sensitivity and the logit of (1-specificity) was 0.286 ($p=0.344$), confirming that the threshold effect is not responsible for the variation in accuracy between studies.

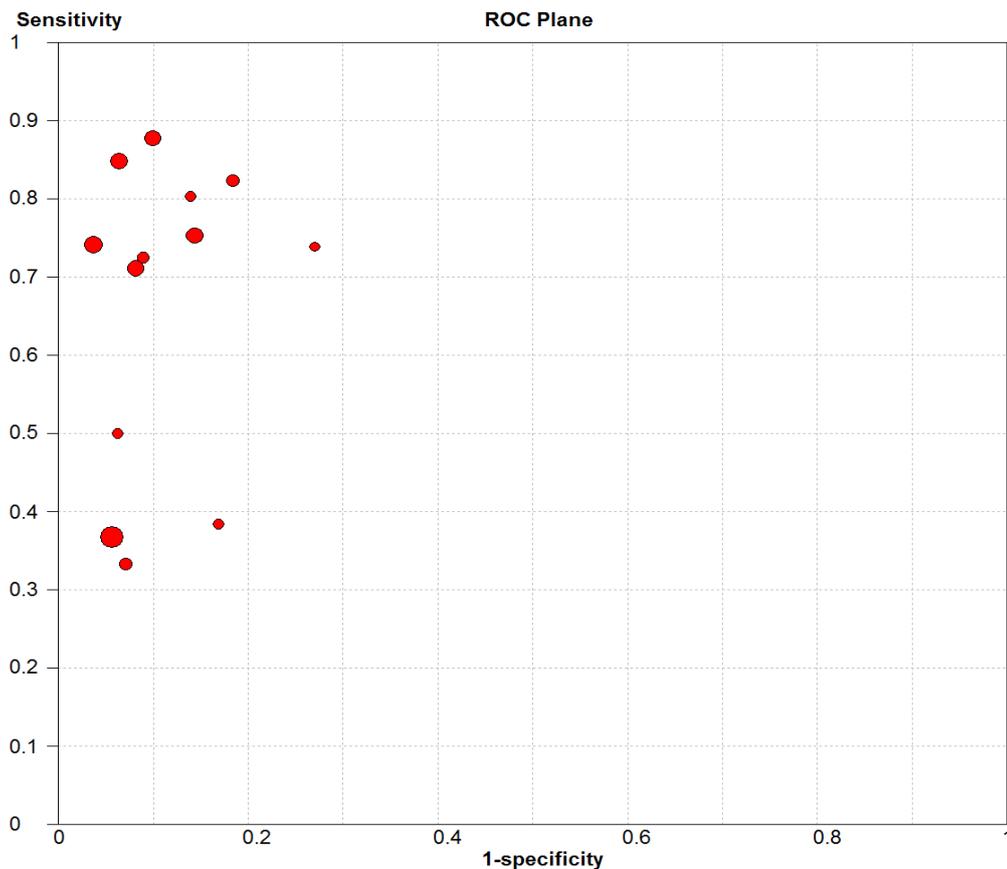


Figure 2-8. Sensitivity and 1-specificity plotted in Receiver Operating Characteristic curve for individual studies and subsets.

Subgroup analysis

Subgroup analysis was based on different study characteristics and perceived sources of bias and applicability uncovered in the QUADAS assessment. Studies at 3T with and without an endorectal coil demonstrated the highest pooled sensitivity of 0.76 (95% CI, 0.71–0.80) and 0.74 (95% CI, 0.71–0.79) respectively. When assessing protocols with a maximum b -value > 1000 sec/mm², the pooled specificity and AUC of the sROC were greater: 0.94 (95% CI: 0.93–0.95) and 0.98,

respectively. A statistically significant improvement was seen using assessment of tumour presence as a visual threshold versus quantitative ROI measurements ($p = 0.03$). The diagnostic performance of the subgroup analysis and p -values of the above-mentioned factors and others are demonstrated in Table 2-8.

Table 2-8. Results of the subgroup analysis.

Study characteristics	No	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	AUC	p-value*
Total	13	0.59 (0.57–0.61)	0.92 (0.91–0.92)	0.92	
b-value (sec/mm ²)					0.31
1000	9	0.55 (0.53–0.58)	0.90 (0.89–0.92)	0.91	
>1000	3	0.72 (0.67–0.76)	0.94 (0.93–0.95)	0.98	
Field strength					0.20
1.5T	6	0.49 (0.46–0.51)	0.90 (0.88–0.91)	0.89	
3T	7	0.74 (0.71–0.79)	0.93 (0.92–0.94)	0.96	
Coil					0.32
With endorectal	4	0.76 (0.71–0.80)	0.86 (0.83–0.89)	0.84	
Without endorectal	9	0.56 (0.53–0.58)	0.93 (0.92–0.93)	0.94	
Reference standard					0.16
Biopsy	3	0.73 (0.63–0.82)	0.86 (0.82–0.89)	0.86	
Prostatectomy	9	0.56 (0.56–0.60)	0.92 (0.91–0.93)	0.94	
Threshold method					0.03
ADC	7	0.64 (0.59–0.69)	0.87 (0.84–0.89)	0.88	
Visual	6	0.58 (0.55–0.60)	0.93 (0.92–0.93)	0.95	
Patient selection bias					0.29
High risk	10	0.58 (0.56–0.60)	0.92 (0.91–0.93)	0.94	
Low risk	3	0.77 (0.68–0.84)	0.83 (0.79–0.87)	0.86	

AUC – Area under the curve. *p-value – comparison of diagnostic odds ratio of subgroups

Discussion

This analysis indicates that high b-value imaging is a good diagnostic tool for detecting prostate cancer. The results of the threshold method subgroup analysis imply that there is a benefit in using a higher maximum b-value in a clinical setting. The lesser value of quantitative ADC thresholding as a tool for detecting tumour is in line with PI-RADS version 2 recommendations (standardized reporting standards for prostate MRI) [67]. The evidence on which this analysis was made was graded as 'good' using the QUADAS-2 tool, however results should be interpreted with caution given the significant statistical heterogeneity in the meta-analysis.

There have been multiple meta-analyses investigating the diagnostic accuracy of DWI alone or in combination with other imaging techniques [100,101,131]. The pooled sensitivity, specificity, and AUC of this study were 0.58, 0.92 and 0.92 respectively, which is similar to Jie's [100] meta-analysis of DWI alone. This is likely due to overlap of included studies, with nine of the ten included studies featured in their meta-analysis. However, in contrast to Jin's meta-analysis of all b-values, the sensitivity was lower in this study (0.58 vs. 0.77), but the pooled specificity and AUC were higher (0.92 vs. 0.84 and 0.92 vs. 0.88 respectively) [131]. This suggests high b-value imaging may help to rule out significant prostate cancer.

There was significant statistical heterogeneity between the included studies demonstrated in the Inconsistency value and Chi-Squared of the Q statistic results. Heterogeneity between studies is also demonstrated in the visual assessment of the sensitivity, specificity, and positive and negative likelihood ratio Forest plots, particularly the former. Following assessment of the ROC curve and Spearman rank correlation coefficient of the logit of sensitivity versus logit of (1-specificity) results it was concluded that the heterogeneity could not be explained by threshold effect.

Heterogeneity in a meta-analysis can come from multiple sources, including both clinical and study parameters. Potential clinical parameter diversity between studies includes patient age, analysis of patients with and without prostate cancer versus just those with cancer, and patient age. Examples of study design variability exist in the method of assessment for tumour, prospective versus retrospective analysis, use of coil or antispasmodics, and choice of b-values analysed.

Results of the studies were pooled in a meta-analysis and given the significant statistical heterogeneity, a random-effects model was used. Diagnostic results of the individual studies were pooled to provide an overall estimate of different diagnostic parameters. This is not recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy as it is felt that pooling studies which used different threshold methods creates a notional average which is not useful clinically [154]. However, despite this advice it is relatively common practice to include pooled results in diagnostic test accuracy meta-analyses [145]. RevMan, the meta-analysis software created by Cochrane has many functions including the creation of Forest plots, but it does not provide pooled results. On the other hand, MetaDisc does provide pooled results for random and fixed effects models [155], but again advises caution in the presence of statistical heterogeneity.

When statistical heterogeneity exists that cannot be explained by threshold effects a subgroup analysis can be performed to look for reasons for the heterogeneity of the results. It is important to identify subgroups *a priori* as pooling of these subsets should only be performed if the subsets are homogenous. Performing subgroup analysis of low numbers of studies can introduce more heterogeneity. In this meta-analysis, the subgroups assessed were chosen prior to analysing the data, however, it was unclear at the outset how many studies would be in each subgroup leading to some with only 3 studies included. The subgroups' pooled results should be assessed with even more caution.

Meta-DiSc provides a measure of difference between covariates within the subgroup analysis. The p-value provided is a measure of the strength of difference between the diagnostic odds ratio of each subgroup. Diagnostic odds ratio is used as it provides an overall marker of accuracy and includes the results of the sensitivity, specificity, and positive and negative likelihood ratios. The diagnostic odds ratios of the individual results were not displayed as the result alone have little meaning clinically and interpreting accuracy as a ratio of odds is challenging. Therefore, the AUC of the sROC was chosen as it is a more well-known indicator of overall accuracy.

Improved tumour contrast at high b-values comes at the cost of decreased signal-to-noise [156]. This can be mitigated through the use of 3T field strength. Most of the diagnostically specific high b-value diagnostic accuracy studies use 3T [83,85,88,157]. The subgroup analysis of field strength demonstrated a trend towards improved accuracy with 3T. Although, the diagnostic odds ratio of the 3T subgroup was not significantly superior to the 1.5T subgroup, the sensitivity and specificity were significantly higher (assessed by reviewing the confidence intervals) and the AUC improved. The sensitivity results of the 3T subgroup alone are similar and the specificity and accuracy are better than those found in Wu's meta-analysis of accuracy of visual assessment of combined T2WI and DWI sequences for prostate cancer detection [101].

An important subgroup analysis was the comparison of the qualitative and quantitative methodologies. They are two very distinct and differing methodologies, both of which have merit. The visual assessment of tumour is more clinically applicable as this is what radiologists use in their daily practice interpreting prostate MRI. ADC measurement and cut-offs do provide diagnostic information, but they are not recommended currently for routine use, particularly as the ADC values of tumours vary with field strength, b-value, and other scanning parameters [94,107]. A further advantage of performing pooling of the results, on these subgroups, are that the studies within each subgroup are more homogenous thus providing more certainty about the

accuracy of the pooled results, although, the methods used in the ADC value measurements studies were varied.

About half of the studies qualitatively assessed the ability of blinded readers to visually detect prostate cancer either by answering a binary question regarding cancer presence or using a probability scale [90,91,141,142]. The remaining studies used ROI-based ADC calculations to determine prostate cancer presence or absence, or used a scale of ADC values to predict cancer [147–152]. The diagnostic performance of tumour visual assessment on ADC maps was significantly better than quantitative ADC methods, with visual assessment giving similar results to Wu's combined T2WI and DWI meta-analysis [101], indicating potential value for high b-value imaging clinically.

The variability amongst the ADC assessment study methodologies was associated with a variability in the diagnostic results of the studies resulting in wide confidence intervals. The pooled results of this subgroup should be interpreted cautiously as it likely does not reflect the usefulness of the technique accurately.

The homogeneity of the qualitative studies' methodology is seen in the blinded visual assessment of images by readers, as well as in the use of radical prostatectomy specimens as reference standard, and finally in the division of the prostate into sectors. By using a sectoral-based qualitative accuracy assessment it increases the number of data points significantly, as each sector is considered a single data point rather than an individual patient or tumour being considered a data point. However, sectoral assessment of the prostate is flawed and not the most applicable clinically. As mentioned previously a lesion-based analysis does not provide an ability to measure true negatives, but the ability to identify individual lesions and stage them accurately is, perhaps, more beneficial to clinicians.

A potential explanation for the relatively poor performance of quantitative ADC methods is that four of the six studies used biopsy as a reference standard whereas all visual assessment studies used radical prostatectomy specimens. TRUS biopsy as a reference standard is limited given its poor octant localization, small sampling volume, and substantial number of tumours missed [158,159]. There was a trend towards improved diagnostic performance in studies which used radical prostatectomy as a reference standard in comparison to biopsy studies, although the difference was not statistically significant.

No other subgroup analyses demonstrated a significant difference between the groups. In some subgroups a statistical difference would have been difficult to demonstrate given the small numbers of studies. For example, only three studies used biopsy as a reference standard, but despite this, this subgroup analysis provided the second strongest statistical source of heterogeneity ($p = 0.18$). The limitations of biopsy as a reference standard are described above, but limiting MRI assessment to patients who have had a prostatectomy introduces patient selection bias, as patients considered for radical prostatectomy tend to be younger, fitter, and have clinically significant tumours, prompting surgery. Radical prostatectomy allows examination of the entire gland including the anterior gland (which TRUS cannot) and detects multifocality, which is frequent [37,160]. Eight of nine studies in the prostatectomy subgroup assessed for multifocal disease, or for tumour in multiple segments of the prostate, and this subgroup's results may be more representative of the diagnostic accuracy of high b-value diffusion.

There are limitations to the review and meta-analysis. Limitation by language and database may have introduced bias. The use of two larger databases, and grey literature, should encompass most eligible English language studies. Publication bias was not assessed. A Deek's funnel plot is considered the most accurate method of assessing accuracy in meta-analyses of diagnostic test accuracy, however in the presence of heterogeneity of diagnostic odds ratio results and with a small numbers of included studies the test is of low power [161]. Finally, this study was restricted

to testing localization of prostate cancer within the gland. This is important in determining the accuracy of high b-value diffusion, but not the only useful outcome. Identifying the presence of capsular breach, seminal vesicle invasion and pelvic lymphadenopathy are important staging and prognostic characteristics not assessed in this meta-analysis.

In conclusion, these findings should be considered cautiously given the degree of statistical heterogeneity. However, this meta-analysis demonstrated that high b-value diffusion is a valuable diagnostic tool, with a sensitivity of 59%, specificity of 92% and sROC AUC of 0.92. There was better diagnostic performance by visual assessment of high b-value DWI studies compared to ADC quantification highlighting the potential value of high b-value DWI clinically, particularly at 3T. The methodologies and patient selection employed by the visual assessment studies are similar to the methods used in the FODIP study, thus providing a marker of accuracy to compare assessed DWI sequences against.

Chapter 3 Meta-analysis: Accuracy of DWI and T2WI

Background

The results of the previous chapter's systematic review and meta-analysis provided a rationale for using higher b-values and suggest it is a good diagnostic test. Furthermore, it added to the evidence that a qualitative assessment of diagnostic performance of DWI for prostate cancer detection is advised over a quantitative analysis of ADC. Its limitations, however, were the small number of studies included in the meta-analysis, particularly, if assessing only the visual assessment subgroup, which had only four studies.

The majority of diagnostic test accuracy studies of DWI in prostate MRI include T2WI sequences in their assessment for tumour. This realisation became apparent to the two reviewers during the study identification step of the high b-value meta-analysis and that, although, the meta-analysis could provide useful information it would not provide much valuable information on comparisons of b-values.

Although, the T2WI sequence is an important sequence in the MRI prostate protocol and part of many diagnostic accuracy studies research protocols it is rarely the focus of the study. Despite the technical parameters of the T2WI sequence usually being described in a study it rarely features as a point of discussion or as an explanation for a study's results. The T2WI sequence is included in these studies as it is an imaging mainstay in the prostate MRI protocol; its benefits, particularly in the transitional zone, cemented in reporting guidelines [67,68]; and, perhaps because it usually is part of research protocols, new studies include it when designing their study to facilitate comparison with the wider literature.

A diagnostic test accuracy meta-analysis of T2WI and DWI sequences for detecting prostate cancer has been performed [101]. This study provided an understanding of the clinical benefit as

it included only studies visually assessing for tumour. The study was published in 2012 and included only 10 primary studies. In addition, they assessed b-value subgroups, comparing b-values of less than 1000 sec/mm² with those 1000 sec/mm² and greater, but found no significant difference in pooled sensitivity and specificity. Like most studies, although including T2WI sequences in their analysis, Wu's meta-analysis did not comment or assess the technical parameters of this sequence. Finally, a potential limitation of this study was it did not perform a QUADAS-2 assessment and did not comment on their adherence to the PRISMA statement.

Repeating the T2WI and DWI diagnostic test accuracy meta-analysis was considered to be beneficial. Firstly, it had been 5 years since this meta-analysis meaning more studies could now be included, which had the potential to provide a more confident understanding of the diagnostic performance. In addition, a subgroup analysis including more studies increases the surety of any potential significant differences between groups. The subgroup analysis possibly could identify the ideal b-value so that it could be implemented when choosing b-values for the FODIP study DWI sequences. Finally, the FODIP study planned to also assess the accuracy of DWI and T2WI together and therefore, this meta-analysis of studies that visually assessed for tumour would provide a reliable benchmark to compare the diagnostic performance of the radiologists in the FODIP study against.

Materials and methods

As in the previous meta-analysis it was registered prospectively with PROSPERO International prospective register of systematic reviews, reference number: 42016036196 [162], and was carried out in accordance to the Preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidance [128].

A systematic review of the literature was independently undertaken by two reviewers (KG and TSy) to identify studies which investigated the diagnostic accuracy of DWI and T2WI MRI in the

detection of prostate cancer. Searches were performed using MEDLINE and EMBASE electronic databases as well as OpenSIGLE to explore sources of unpublished grey literature. The Science Citation Index was used to identify articles which cite those identified with the original search terms. Once eligible studies were found their reference lists were manually searched for further potential papers. The search strategy for MEDLINE including Boolean operators and MeSH terms is presented in Table 3-1, the same search strategy was used for each database with alterations to suit. All studies were included up to the date of the search; 1st September 2017.

Table 3-1. MEDLINE search terms and strategy.

1	Exp Prostate* Neoplasm*/
2	Prostat* cancer*.mp.
3	Prostat* carcinoma*.mp.
4	Or/1-3
5	Exp Diffusion magnetic resonance/
6	DW magnetic resonance imaging.mp.
7	DWI.mp.
8	DW-MRI.mp.
9	Or/5-6
10	4 and 9
11	10 limit to Human studies
12	11 limit to English Language

Eligibility

The eligibility criteria for the studies included within the systematic review were

- Both DWI and T2WI MRI used in combination for the visual assessment of prostate cancer.
- Assessment of the pre-treatment patient population
- Histopathological reference standard, be that biopsy or radical prostatectomy.
- Sufficient information to produce a 2x2 contingency table (true positives, false positives, false negatives and true negatives) to calculate sensitivity and specificity.
- Published in English
- Assess more than 10 individual patients.

Studies would be excluded if they did not satisfy the inclusion criteria above or they used a different combination of imaging sequences from DWI and T2WI so that individual data for the desired combination could not be extracted. They would also be excluded if an ADC cut-off was used to discriminate malignant from benign tissue as opposed to visual assessment by radiologists. Studies were not excluded by country of origin, age of patients or study design.

Study identification

Initially papers were reviewed by relevancy of title and then abstract. Residual articles had their full text reviewed against the inclusion and exclusion criteria. This was also performed independently by the same two reviewers; any disagreement was solved by consensus or a third expert reviewer if necessary (TSm).

Data extraction

The following data were extracted from each eligible study; year of publication, country of origin, patient group, number of patients, average age and prostate specific antigen (PSA), study design

(retrospective or prospective) and the histopathological reference standard used. Further information on the imaging specifications was also gathered; field strength, coil used, field-of-view, b -value, and whether they used DWI, ADC maps or both to assess each patient. True positives, false positives, false negatives and true negatives were also extracted for pooling results, when insufficient data was available reviewers would manually calculate them from other reported statistics when possible. All data extraction was independently verified by two reviewers (KG, TSy).

Quality assessment

The quality of the each included study's methodology was assessed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [122]. This was also undertaken independently by two reviewers (KG, TSy) and disagreement resolved with consensual discussion and consulting a third expert reviewer if consensus could not be met (TSm).

Statistical analysis

The sensitivity and specificity with 95% Confidence Intervals were calculated for each study using the extracted details of the 2x2 contingency tables, and Forest plots produced.

Initially heterogeneity of studies was examined visually using the data extraction tables, followed by statistical analysis using the inconsistency value (I^2) and Q statistics of the Chi-squared value. An $I^2 > 50\%$ and p -value < 0.10 respectively would indicate significant statistical heterogeneity and a random-effects model would be applied to data pooling. Pooled results for sensitivity, specificity, positive and negative likelihood ratios with 95% Confidence Intervals and a sROC curve was also presented.

To explore predictable sources of heterogeneity between included studies, the sensitivity and 1-specificity was plotted on an ROC plane to visually assess the presence or absence of a 'shoulder arm' shape, which would indicate a threshold effect. This was also tested statistically with the

spearman correlation coefficient of the logit of sensitivity and logit of (1-specificity) with a p -value <0.05 suggesting a threshold effect. Subgroup analysis was performed for; b -values ($<1,000$, $1,000$ and $>1,000$), field strength (1.5T and 3T), coil type (endo-rectal and body), method of assessment (DWI, ADC or both), reference standard (Biopsy and radical prostatectomy), tumour zone (Peripheral or transitional zone) and study design (retrospective and prospective). Pooled sensitivities, specificities, positive and negative likelihood ratios and meta-regression of diagnostic odds ratio with p -values was performed for these sub-groups. All statistical analysis was conducted on Meta-DiSc (version 1.4, Javier Zamora). The choices of statistical tests used are identical to those outlined in Chapter 2 (Page 61).

Results

Search results

With the presented search strategy 2,825 citations were discovered, after duplicates were removed there was 1,880 unique articles. 33 studies were included in the final analysis after reviewing against the eligibility criteria. The PRISMA flowchart of the search results is presented in Figure 3-1.

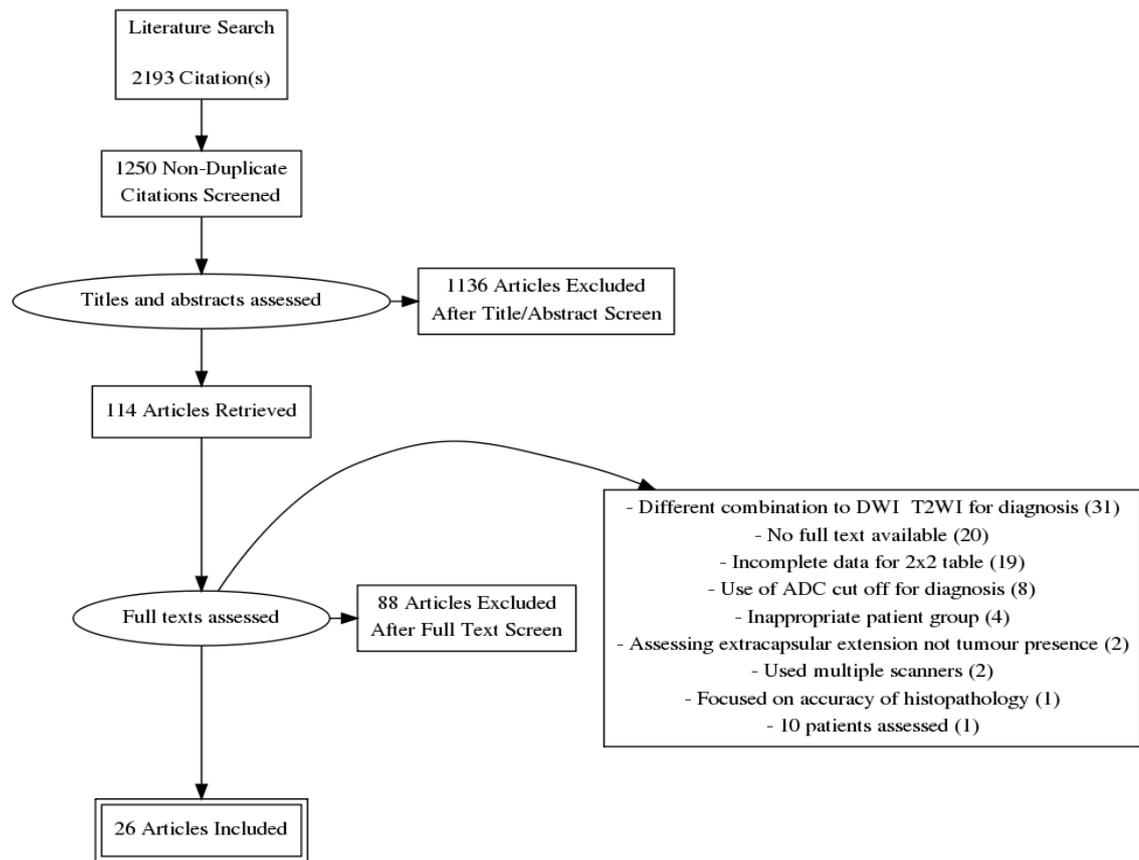


Figure 3-1. PRISMA Flow Diagram.

Quality assessment

The full results of the QUADAS-2 appraisal is presented in Figure 3-2. The strengths across the included studies were that a majority used consecutive patient selection with appropriate inclusion and exclusion criteria, however, two studies [163,164] limited their investigation to transitional zone tumours and another [165] to patients with ‘low risk’ cancer. Therefore, a subgroup analysis was deemed particularly important to assess the differences between peripheral and transitional zone tumours. Another strength was that all index tests used were applicable to clinical practice with no peculiar imaging methods outside of the norm. All but one study imaged patients after a positive biopsy while patients in Tanimoto et al. had a pre-biopsy MRI [166]. A number of studies did not state the timings between biopsy and MRI [139,141,167,168], which could have implications if the timing was long enough to cause a disparity between the images and histopathology correlation, or too short resulting in an increased incidence of post-biopsy

haemorrhage potentially affecting image interpretation. Kitajima et al. [138] and Morgan et al. [169] reported delays between biopsy and imaging much less than the recommended 6 weeks [68].

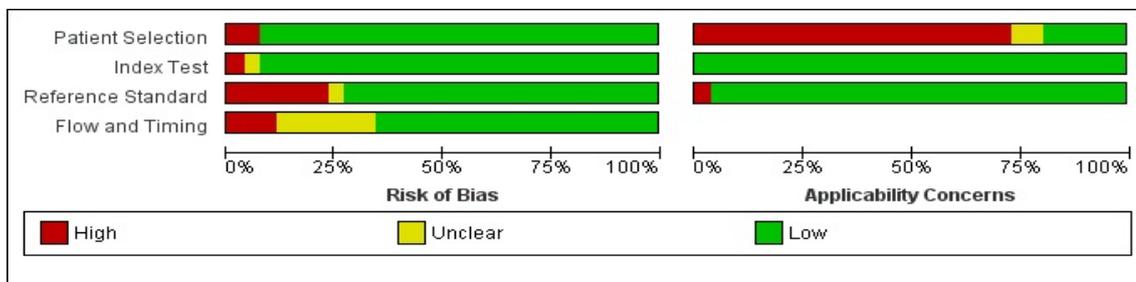


Figure 3-2. QUADAS-2 results summarising the proportion of low, high and unclear risk of bias and applicability concerns.

The predominant weakness of included studies was the applicability of the patient groups selected by the studies, as they were often limited to those who underwent radical prostatectomy, who in general would be younger with a narrower range of tumour staging. This is acceptable to obtain a reference test with low bias.

Study characteristics

The data extracted for study characteristics are described in Table 3-2 – Table 3-4. There were 2,949 patients across the 33 studies. The mean age and PSA was 65.1 (range=41-86) years and 9.0 (range=0.4-130) ng/mL respectively. A majority of studies (n=20) used a retrospective study design as opposed to prospective (n=13). Most of the studies (n=19) used 3T field strength, thirteen used 1.5 T, and one used both 1.5 and 3T. The highest b-value used by a study ranged from 600 – 2000 sec/mm². The majority of studies (n=18) used a maximum b-value of 1,000 sec/mm², and all studies using b-values over 1,000 used 2,000 sec/mm². Nine studies used DWI images alone for diagnosis, while seven used ADC maps only, and seventeen using both. Most (n=20) used radical prostatectomy as reference standard. Seven used TRUS biopsy, two used MRI guided biopsy, and one used trans-perineal biopsy and another used a mixture of TRUS biopsy and radical prostatectomies.

Table 3-2. Principle characteristics of included studies.

Study	Year	Country	No. of patients	Mean Age (range)	Mean PSA (Range)	Design
Agha [170]	2015	Egypt	20	n/a	n/a	Pro
Bains [171]	2014	Switzerland	111	64*(43-82)	n/a (0.7-112.2)	Pro
Baur [172]	2016	Germany	44	66 (46-81)	12.3 (5.2-70)	Pro
Brendle [119]	2016	Germany	15	66 (52-76)	11.8 (3.3-65.4)	Pro
Costa [173]	2016	USA	49	63 (49-79)	11.2 (2.5-48.5)	Pro
Haider [140]	2007	Canada	49	61*(46-75)	5.4* (0.9-26)	Pro
Hoeks [163]	2013	Netherlands	28	n/a (45-73)	n/a (1.9-44)	Retro
Isebaert [141]	2013	Belgium	75	66*(49-64)	10.4 (1.5-70.9)	Pro
Iwazawa [174]	2011	Japan	178	69 (41-86)	n/a	Retro
Jung [164]	2013	South Korea	156	59*(42-75)	4.9 (0.4-93.7)	Retro
Katahira [86]	2011	Japan	201	69 (43-80)	13.2 (2.6-114)	Retro
Kim [165]	2014	South Korea	100	63* (51-76)	6.5* (2.2-9.5)	Retro
Kitajima [138]	2010	Japan	53	69* (56-84)	11.1*(4.2-112)	Retro
Kuhl [175]	2017	Germany	542	64.8 (42-80)	8.5 (3.2-67.5)	Pro
Lim [142]	2009	South Korea	52	65 (48-76)	10.5 (1.2-79.6)	Retro
Loggitsi [176]	2017	Greece	26	63.7 (48-73)	8.1 (2-21.9)	Pro
Morgan [169]	2007	United Kingdom	54	68 (52-80)	10 (n/a)	Pro
Ohgiya [85]	2012	Japan	73	70 (n/a)	11.7* (n/a)	Retro
Petrillo [167]	2014	Italy	136	66 (n/a)	6.8 (n/a)	Pro
Rosenkrantz 2011[177]	2011	USA	42	62 (47-76)	6.2 (1.3-32.5)	Retro
Rosenkrantz 2015[178]	2015	USA	106	62 (56-81)	6.9 (n/a)	Retro
Shimofusa [179]	2005	Japan	37	71 (54-82)	21.8 (4.5-130)	Retro
Shinmoto [180]	2015	Japan	87	n/a (51-75)	n/a (2.8-35.2)	Retro
Stanzione [181]	2016	Italy	82	65 (n/a)	8.8 (n/a)	Pro
Tanimoto [166]	2007	Japan	83	67 (53-87)	19.4 (n/a)	Pro
Thestrup [182]	2016	Denmark	204	64.1 (45-75)	14 (2.2-120)	Retro
Ueno 2013 [83]	2013	Japan	73	67 (50-77)	9.51 (2.9-49)	Retro
Ueno 2013[168]	2013	Japan	80	67 (50-77)	9.51 (2.9-49)	Retro
Ueno 2015[139]	2015	Japan	31	65 (51-81)	8.6 (4.7-16.5)	Retro
Vargas [104]	2011	USA	51	56* (46-74)	5.3 (0.4-62.2)	Retro
Yoshimitsu [183]	2008	Japan	37	66 (56-75)	11.9 (0.7-54.8)	Retro
Yoshizako [184]	2008	Japan	23	65* (52-76)	n/a	Retro

*Median, n/a - not available, PSA - Prostate specific antigen (ng/mL), Pro – prospective, Retro – retrospective

Table 3-3. Imaging and methodological characteristic of included studies.

Study	Field Strength	ERC	FOV (cm)	<i>b</i> -value (sec/mm ²)	Reference	AS	Method
Agha	3T	N	30x30	0, 1000	Bx	U	Both
Bains	3T	N	n/a	0, 500, 1000	RP	Y	Both
Baur	3T	Both	20x20	0, 100, 500, 1000	MR	N	Both
Brendle	3T	N	27.6x28	50, 800	RP	U	Both
Costa	3T	Both	16x16	0-2000	Mix	U	Both
Doo	3T	N	28x28	0, 1000	RP	U	ADC
Haider	1.5T	Y	14x14	0, 600	RP	U	ADC
Hoeks	3T	Y	20.4x20.4	0, 50, 500, 800	RP	U	Both
Isebaert	1.5T	N	30.9x38	0,50,100,500,1000	RP	U	DWI
Iwazawa	1.5T	N	30x30	0, 1000	Bx	U	DWI
Jung	1.5T/ 3T	Y	12x12/14x14	0, 1000	RP	U	ADC
Katahira	1.5T	N	35x35	0,1000, 2000	RP	U	DWI
Kim	3T	N	34x16.8	0, 100, 1000	RP	Y	Both
Kitajima	3T	N	35x25	0, 1000	TP	N	Both
Kuhl	3T	N	21x21	0, 800, 1000, 1400	MR	U	Both
Lim	1.5T	Y	22x22	0, 1000	RP	Y	ADC
Loggitsi	1.5T	N	10x10	0, 250, 500, 750, 1000	RP	U	Both
Morgan	1.5T	Y	20x20	0, 300, 500, 800	Bx*	Y	ADC
Ohgiya	3T	N	35x35	0,500,1000,2000	Bx	U	DWI
Petrillo	1.5T	Y	13.6x16	0,50,100,150,300,600, 800	Bx	N	Both
Rosenkrantz 2011	1.5T	N	30x24.4	0, 500, 1000	RP	U	DWI
Rosenkrantz 2015	3T	N	20x20/ 28x22	50,1000,2000	RP	U	Both
Shimofusa	1.5T	N	20x20	0, 1000	Mix	U	DWI
Shinmoto	3T	N	24x24	0, 1000	RP	Y	ADC
Stanzione	3T	N	20x20	0, 400, 2000	Bx	U	Both
Tanimoto	1.5T	N	36x36	0, 1000	Bx	U	Both
Thestrup	3T	U	19x19	0, 100, 800, 2000	Mix	Y	Both
Ueno 2013	3T	N	45x45	0, 1000, 2000	RP	Y	DWI
Ueno 2013	3T	N	n/a	0, 1000, 2000	RP	Y	DWI
Ueno 2015	3T	N	45x36	0, 2000	RP	Y	DWI
Vargas	3T	Y	14x14	0, 700/ 0, 1000	RP	U	ADC
Yoshimitsu	1.5T	N	24x24	0, 500, 1000	RP	U	Both
Yoshizako	1.5T	N	42x21	0, 1000	RP	Y	Both

T - tesla, ERC – endo-rectal coil, Y – Yes, N – No, FOV - field-of-view, Bx – biopsy, * trans-perineal biopsy, RP - radical prostatectomy, Mix - mixture of Bx and RP, AS - antispasmodic, U – unclear

Table 3-4. Diagnostic Performance of included studies.

Study	TP	FP	FN	TN	Sens	Spec	Notes
Agha	10	1	5	4	0.67	0.80	
Bains	73	7	7	24	0.91	0.77	
Baur	14	11	0	18	0.97	0.62	Body coil
	10	7	1	21	0.91	0.75	Endo coil
Brendle	17	2	12	149	0.59	0.99	
Costa	20	19	06	73	0.44	0.79	Body coil
	76	51	22	145	0.78	0.74	Endo- coil
Doo	113	21	58	216	0.66	0.91	
Haider	120	39	29	204	0.81	0.84	
Hoeks	65	39	47	101	0.58	0.72	TZ
Isebaert	444	79	546	731	0.45	0.90	
Iwazawa	238	223	80	883	0.75	0.80	
Jung	91	62	84	699	0.52	0.92	TZ
Kitajima	971	559	616	2669	0.61	0.83	<i>b</i> =1000
	1162	332	425	2896	0.73	0.90	<i>b</i> =2000
Kim	17	7	22	72	0.44	0.91	
Kitajima	75	19	24	306	0.76	0.94	
Kuhl	138	49	9	346	0.94	0.88	
Lim	199	49	28	348	0.88	0.88	
Loggitsi	43	33	62	330	0.41	0.91	
Morgan	64	56	78	126	0.45	0.69	
Ohgiya	25	5	30	13	0.45	0.72	<i>b</i> =500
	43	4	12	14	0.78	0.78	<i>b</i> =1000
	42	2	13	16	0.76	0.89	<i>b</i> =2000
Petrillo	18	48	7	63	0.72	0.57	
Rosenkrantz 2011	61	29	59	103	0.51	0.78	
Rosenkrantz 2015	34	13	28	561	0.55	0.98	<i>b</i> =1000
	46	10	16	564	0.74	0.98	<i>b</i> =2000
Shimofusa	96	11	15	56	0.86	0.84	
Shinmoto	93	12	58	185	0.62	0.94	
Stanzione	29	1	5	52	0.85	0.98	
Tanimoto	37	6	7	33	0.84	0.85	
Thestrup	65	116	3	20	0.96	0.15	
Ueno 2013	258	87	83	156	0.76	0.64	<i>b</i> =1000
	276	79	65	164	0.81	0.68	<i>b</i> =2000
Ueno 2013	270	119	57	194	0.83	0.62	<i>b</i> =1000
	275	105	52	208	0.84	0.66	<i>b</i> =2000
	272	95	55	218	0.83	0.70	<i>b</i> = <i>c</i> 2000
Ueno 2015	101	63	20	64	0.83	0.50	<i>b</i> =2000
	86	51	35	76	0.71	0.60	<i>b</i> = <i>c</i> 2000

Vargas	65	10	42	157	0.61	0.94	
Yoshimitsu	105	29	42	46	0.71	0.61	
Yoshizako	21	2	5	14	0.81	0.88	TZ

(TP - true positive, TN - true negative, FN - false negative, FP - false positive, sens – sensitivity, spec – specificity, *b* - *b*-value, *c* – computed, PZ - peripheral zone, TZ - transitional zone)

Meta-analysis

The visual assessment of the data extraction tables was deemed homogeneous enough to undertake a meta-analysis with pooling. The pooled sensitivity, specificity, positive and negative likelihood ratio, AUC and diagnostic odds ratio with the Inconsistency values are present in Table 3-5, and Figure 3-3 – 3.6. The pooled sensitivity and specificity of all included studies was 0.69 (95% CI, 0.68–0.69) and 0.84 (95% CI, 0.83–0.85), respectively. While the pooled diagnostic odds ratio (DOR) was 12.27 (95% CI 9.60–15.68). The sROC (Figure 3-7) gave an AUC of 0.839, indicating good diagnostic accuracy [153].

Table 3-5. Sub-group analysis and meta-regression.

Group (number of studies)	Sensitivity (95% CI)	Specificity (95% CI)	DOR	<i>p</i> -value*
Total (n=40)	0.69 (0.68–0.69)	0.84 (0.83–0.85)	12.27 (9.60–15.68)	
<i>b</i> -value (sec/mm ²)				0.068
<1,000 (n=8)	0.60 (0.56–0.64)	0.80 (0.78–0.83)	8.02 (3.18–20.26)	
1,000 (n=26)	0.64 (0.62–0.65)	0.85 (0.84–0.85)	12.56 (9.56–16.50)	
>1,000 (n=15)	0.78 (0.76–0.79)	0.83 (0.82–0.84)	14.32 (9.06–22.65)	
Field Strength				0.418
1.5T (n=18)	0.64 (0.63–0.65)	0.85 (0.84–0.86)	10.68 (7.34–15.55)	
3T (n=18)	0.81 (0.79–0.82)	0.81 (0.79–0.82)	13.77 (9.54–19.88)	
Coil				0.462
Body (n=38)	0.68 (0.67–0.69)	0.85 (0.84–0.85)	13.06 (10.05–16.97)	
Endo-rectal (n=10)	0.68 (0.65–0.69)	0.84 (0.84–0.82)	10.40 (4.87–22.24)	
Tumour zone				0.239
PZ (n=6)	0.71 (0.70–0.73)	0.84 (0.82–0.85)	12.64 (7.13–22.41)	
TZ (n=11)	0.66 (0.64–0.68)	0.88 (0.87–0.88)	13.46 (8.08–22.44)	
Assessment method				0.070
DWI (n=21)	0.68 (0.67–0.69)	0.82 (0.81–0.83)	8.91 (6.80–11.68)	
ADC map (n=8)	0.66 (0.64–0.69)	0.89 (0.87–0.90)	15.44 (6.80–35.05)	
Both (n=20)	0.72 (0.70–0.75)	0.86 (0.85–0.87)	18.58 (9.77–35.30)	
Design				0.918
Prospective (n=16)	0.59 (0.56–0.61)	0.81 (0.80–0.83)	11.93 (6.61–21.54)	
Retrospective (n=33)	0.71 (0.70–0.72)	0.84 (0.84–0.85)	12.56 (9.57–16.49)	
Reference standard				0.420
RP (n=31)	0.67 (0.66–0.68)	0.85 (0.85–0.86)	12.09 (9.24–15.81)	
Biopsy (n=14)	0.73 (0.70–0.76)	0.81 (0.80–0.83)	15.83 (7.27–34.44)	

* Comparison between the diagnostic odd ratio of subgroups, RP – radical prostatectomy

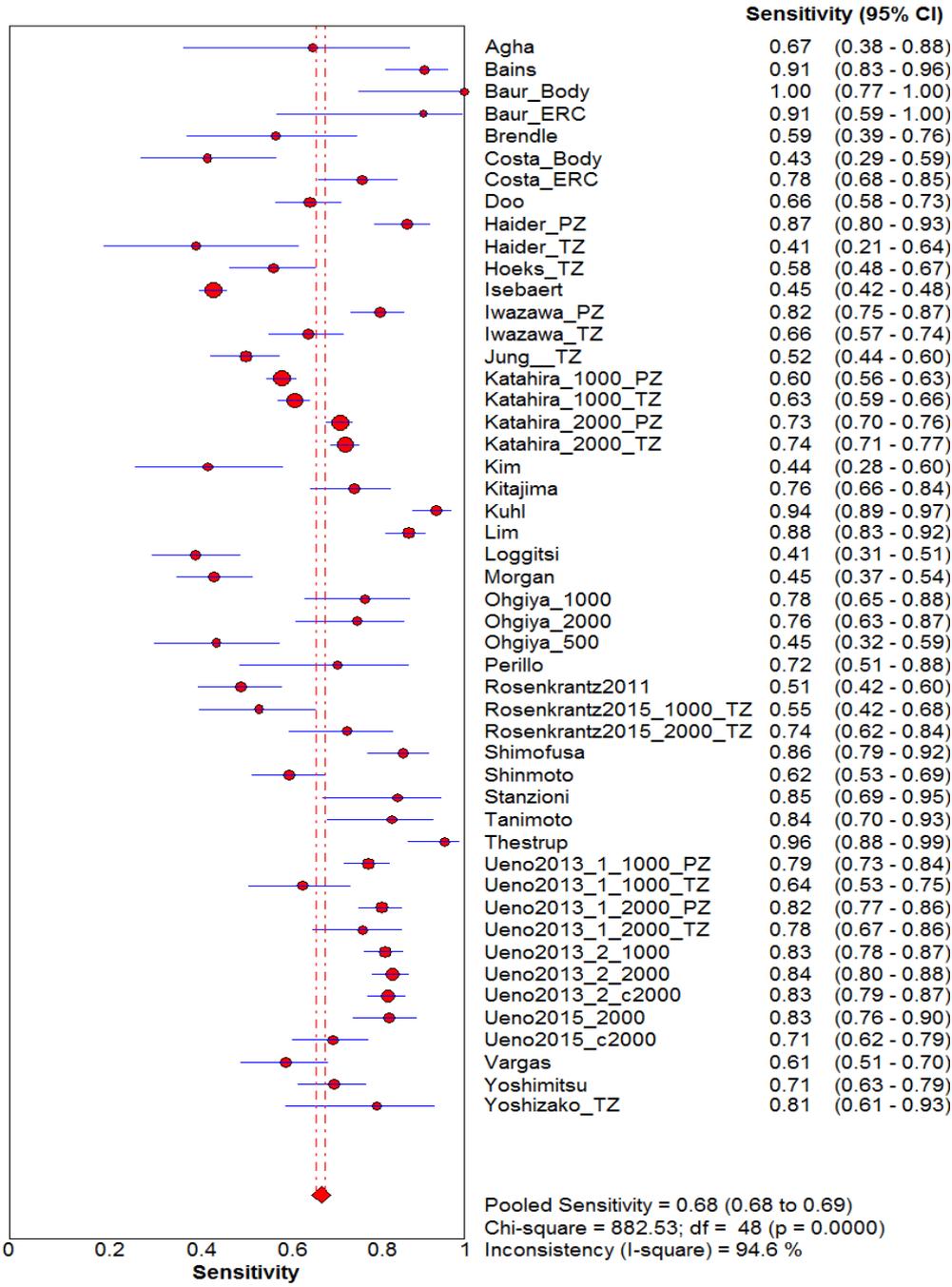


Figure 3-3. Forest plot of sensitivity for detecting prostate cancer including 95% CI, I² value and Q statistic.

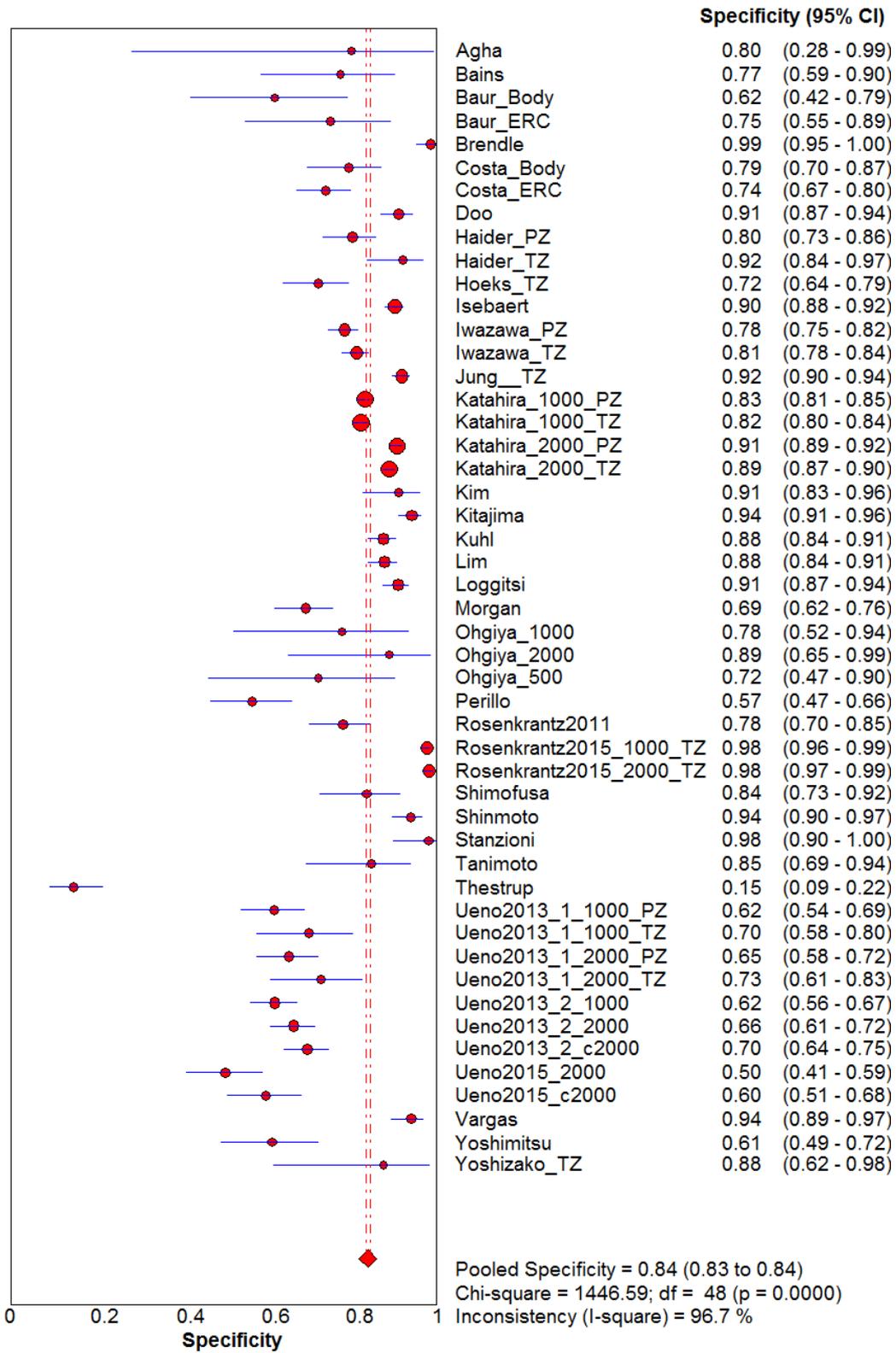


Figure 3-4. Forest plot of specificity for detecting prostate cancer including 95% CI, I² value and Q statistic.

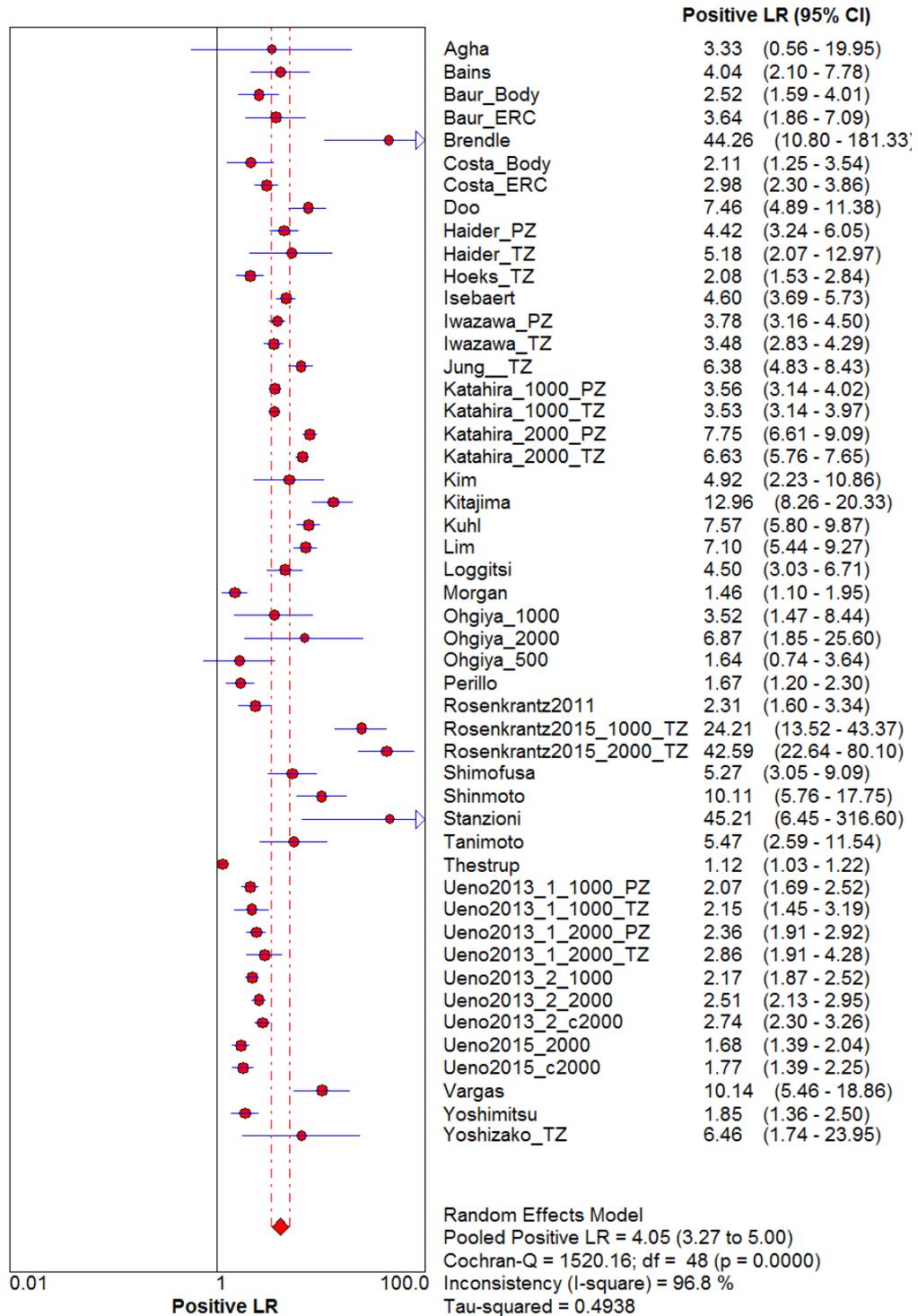


Figure 3-5. Forest plot of positive likelihood ratio for detecting prostate cancer including 95% CI, I² value and Q statistic.

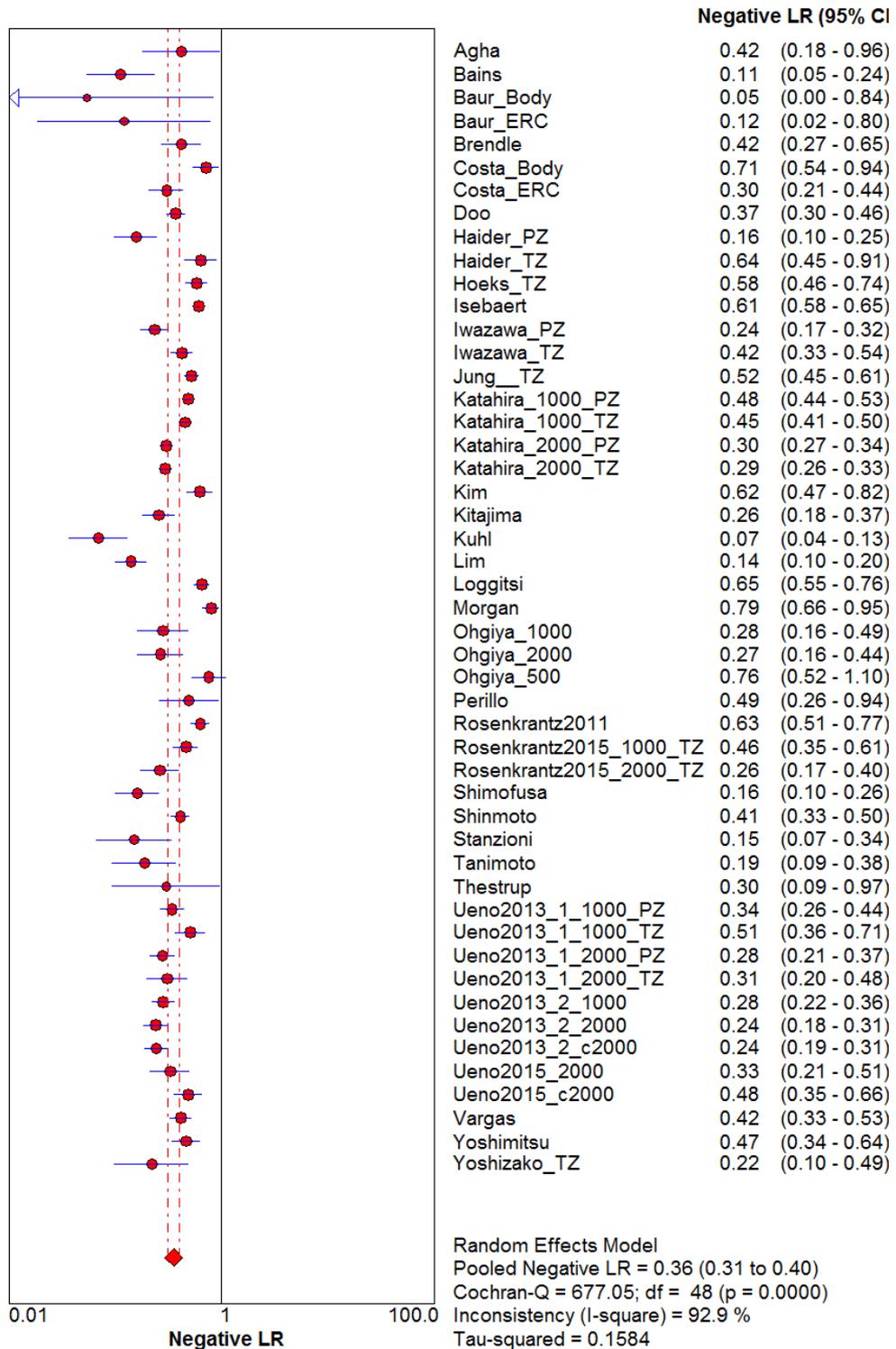


Figure 3-6. Forest plot of negative likelihood ratio for detecting prostate cancer including 95% CI, I² value and Q statistic.

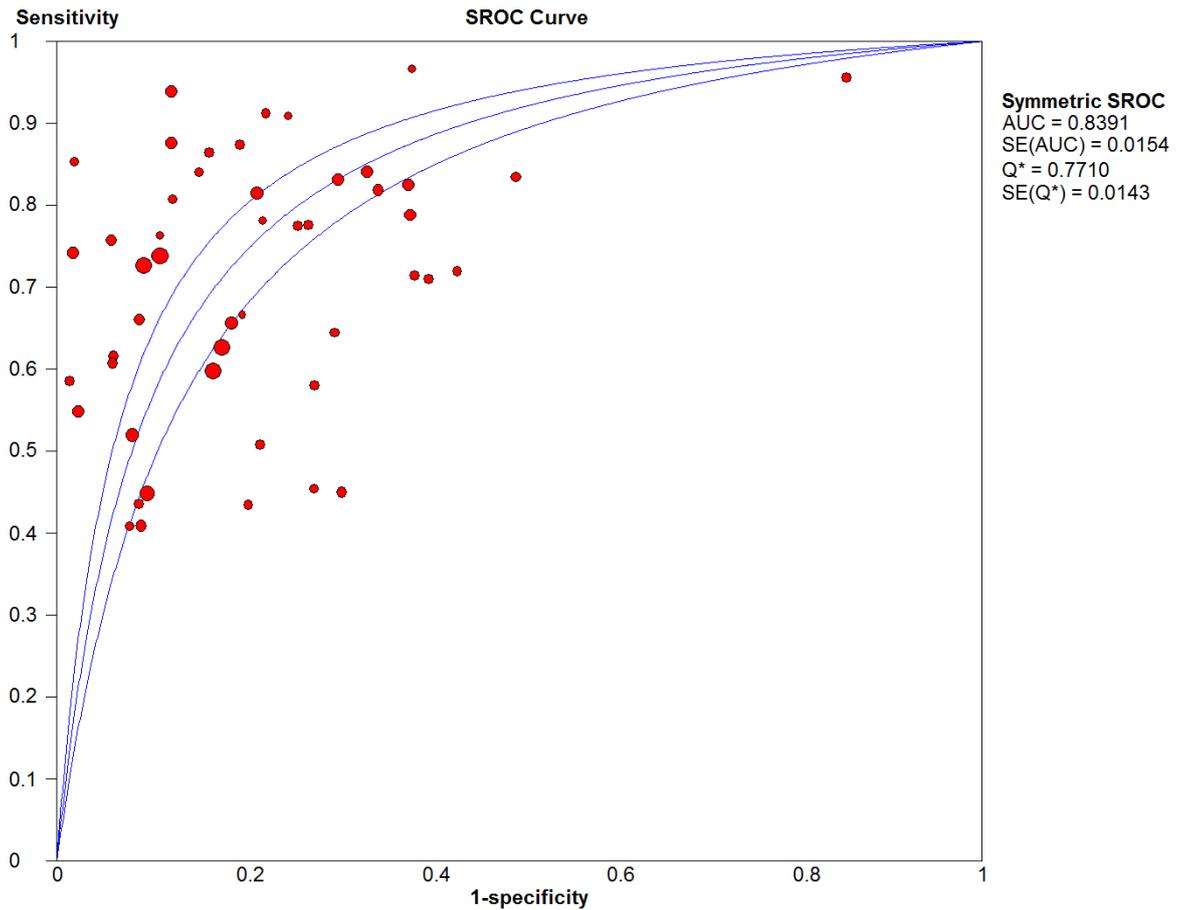


Figure 3-7. Summary receiver operating characteristic (SROC) curve for the detection of prostate cancer.

The I^2 value and Chi-square Q was 94.6% and 882.53 ($p < 0.001$), respectively, for sensitivity and 96.7% and 1446.59 ($p < 0.001$) for specificity indicating a significant level of statistical heterogeneity. The ROC plane (Figure 3-8) did not show a 'shoulder-arm' shape and the spearman rank correlation coefficient of the logit of sensitivity against logit of (1-specificity) was 0.335 ($p = 0.018$), indicating there could be heterogeneity due to a threshold effect.

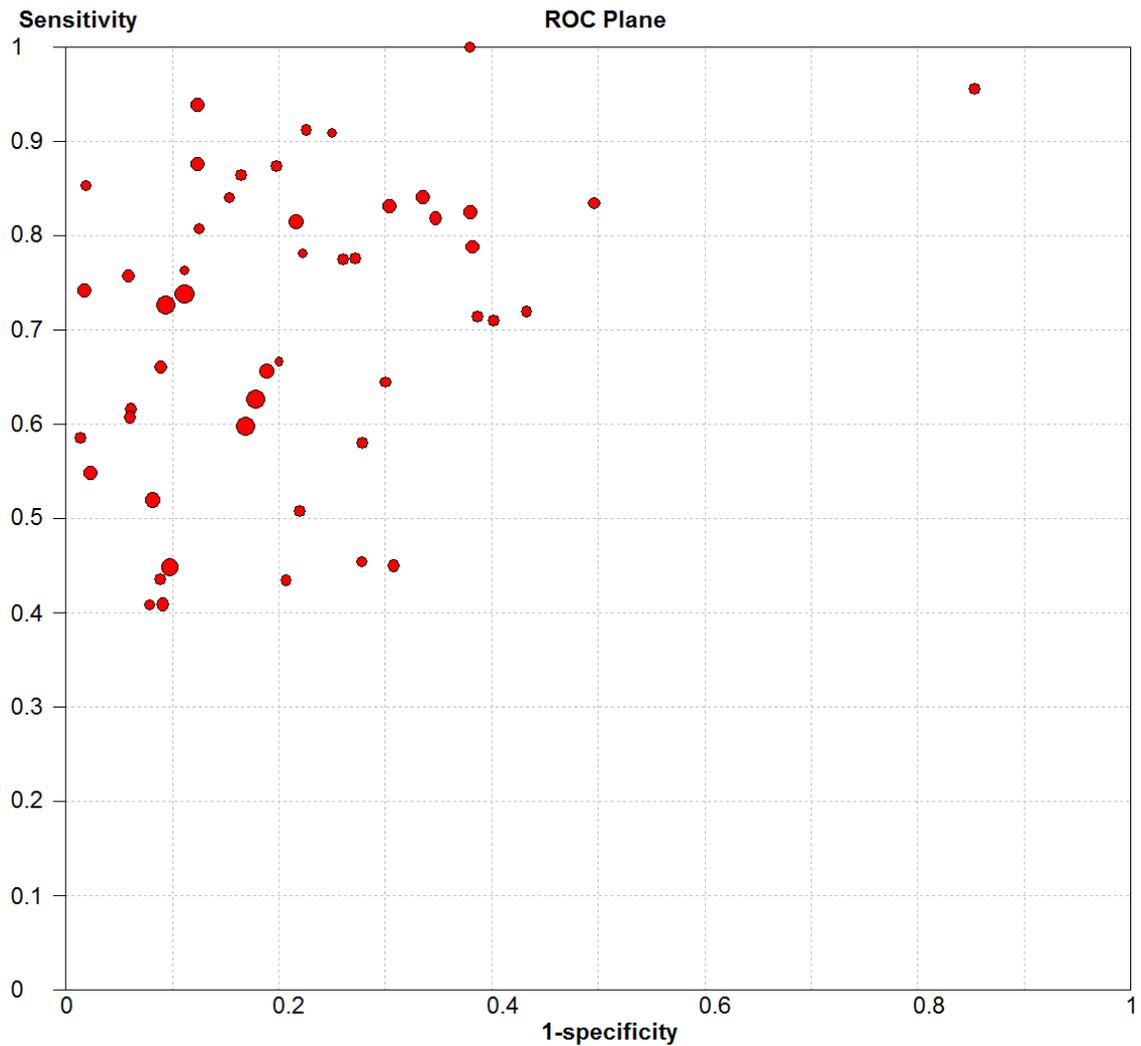


Figure 3-8. Sensitivity and 1-specificity plotted in Receiver Operating Characteristic curve for individual studies and subsets.

Sub-group analysis

The highest DORs were when using ADC maps alone or with DWI for tumour assessment and for b -values greater than $1,000 \text{ sec/mm}^2$. There was a significantly higher sensitivity achieved when using b -values greater than $1,000 \text{ sec/mm}^2$, a 3T field strength, assessing peripheral zone tumours, studies with a retrospective design, and those using biopsy as a reference standard (by assessing confidence intervals). Specificity improved significantly with a 1.5 T field strength, assessing transitional zone tumours, using ADC maps alone and with DWI, and those studies using radical prostatectomy as the reference standard. The complete sub group analysis is shown in Table 3-5.

Discussion

The findings from this systematic review and meta-analysis demonstrate that the diagnostic accuracy of diffusion and T2 weighted imaging of prostate cancer is good, when using a visual assessment [153]. The greatest diagnostic accuracy can be achieved with *b*-values over 1,000 sec/mm² and when assessing lesions with ADC alone and with DWI images. The interplay between sensitivity and specificity can be significantly altered by the choice of field strength and whether tumours originate from the peripheral or transitional zone. The overall strength of the evidence, on which this analysis was based, was graded as good using the QUADAS-2 critical appraisal tool [122]. There was a high degree of unknown statistical heterogeneity so care should be taken when interpreting these results and even though the meta-analysis cannot specify an optimal imaging protocol it does highlight important factors to be considered.

The pooled results match that of Wu and Tan's diagnostic test accuracy meta-analyses of T2WI and DWI, this is likely to be because of the large overlap of included studies [101,185]. Compared to Jie et al., who analysed DWI use alone, there is a higher sensitivity but lower specificity [100]. In comparison to the high *b*-value DWI meta-analysis in the previous chapter the use of T2WI and DWI combined resulted in significantly greater sensitivity, but inferior specificity and AUC when compared to the use of high *b*-value alone. The results of the subgroup analysis of peripheral zone tumours alone are similar to the high *b*-value meta-analysis' pooled results (page 75). This suggests that the addition of T₂-weighted imaging may improve the sensitivity for diagnosing transitional zone tumours, although Jie et al. did not present a subgroup for TZ tumours and it was not possible to extract comparative TZ data in the high *b*-value meta-analysis. This finding supports the present consensus that T2WI should be the predominant imaging technique and DWI a secondary sequence for diagnosing transitional zone tumours [67].

When comparing the visual assessment subgroup of high *b*-value DWI in the previous chapter's meta-analysis (sensitivity – 0.58, specificity – 0.93, AUC – 0.95) with the total pooled results and

matching b -value subgroups ($\geq 1000 \text{ sec/mm}^2$) within this meta-analysis, the sensitivity is significantly worse when visually assessing high b -value images alone, but the specificity and AUC are significantly improved.

There is a significant increase in sensitivity using a b -value greater than $1,000 \text{ sec/mm}^2$ in both meta-analyses and improved specificity with a b -value of 1000 sec/mm^2 and greater when visually assessing with T2WI and DWI. The improved contrast possible with higher b -values by further suppression of normal prostate tissue would explain the increase in sensitivity. By making tumours more visually apparent and the suppression of normal prostate signal results in greater reader confidence in deciding there is no tumour in a region of the prostate.

Two of the studies [139,168] also used visual assessment of computed high b -value DWI images, which in both showed a decrease in sensitivity and increased specificity compared to the natural b -value equivalent. There has been limited research directly comparing the diagnostic accuracy of computed b -values but it shows promise with improved distortion, ghosting and tumour conspicuity [136,186].

A further limitation of the $b > 1000$ subgroup is the small number of included studies. Only 6 were identified, three of which are by the same group [83,168,187] and in at least two of those the same patient cohort was used. This leaves only three studies which reduces the strength of evidence of the subgroup. Furthermore, all studies in this subgroups, except Kuhl's study [175], used a b -values of $2,000 \text{ sec/mm}^2$. Wang et al. and Metens et al. found greatest contrast and image quality using b -values of $1,500$ compared to either $1,000$ or $2,000 \text{ sec/mm}^2$, however there is no apparent data on the diagnostic accuracy of $b 1,500 \text{ sec/mm}^2$ DWI [105,107].

As well as maximum b -value, the minimum number of b -values used to calculate ADC maps has shown to significantly alter ADC values, but there is little evidence about their impact on diagnostic accuracy with visual assessment [88,107]. All but one of the included studies in this

analysis used $b=0$ as the minimum b -value but the number of b -values ranged from two to seven. Thörmer et al. found an improved qualitative image score using just two b -values and a minimum of 50 as opposed to 0 [105], but they tested only a limited number of combinations with a maximum b -value of just 800. Further work directly comparing high b -value ADC maps with different minimum and number of intermediary b -values for visual diagnostic performance would help further optimise mpMRI.

There was an insignificant difference in DOR between 1.5T and 3T studies ($p=0.825$), but there was a significantly higher sensitivity and lower specificity with 3T studies (comparison of confidence intervals). Higher field strengths have the advantage of increased signal-to-noise ratio which gives the potential for better spatial and temporal resolution. Theoretically, the higher field strength may increase the frequency of susceptibility artefact and signal heterogeneity, but there is conflicting evidence to the advantage of 3T over 1.5 T [156].

For some of the studies it was possible to separate the results for peripheral and transition zones. There was a significantly higher sensitivity for the peripheral zone but higher specificity for the transition zone. Transition zone tumours are often of a lower grade than those found in the peripheral zone so may be less apparent on imaging [188,189]. There is also some difficulty in differentiating malignant from benign nodules common in the transitional zone, particularly with DWI. Benign nodules are often heterogeneous and can demonstrate restricted diffusion. Peripheral zone tissue is usually more homogeneous and higher signal than tumour on T2WI. This could explain the difference in sensitivity between the peripheral and transitional zones. One would expect the greater overlap between imaging features of benign and malignant transitional zone lesions would result in greater false positives, thus reducing specificity. In addition, the relative rarity of TZ tumours may explain the drop in sensitivity however the overall DOR was not significantly different. It may be that different imaging parameters are needed for optimal diagnosis of peripheral or transitional zone disease. Given the different challenges of diagnosing

tumours in different regions of the prostate and the different accuracy and strengths of T2WI and DWI, it is important when assessing the diagnostic accuracy of prostate sequences the whole gland and also different zone are assessed. Furthermore, when assessing the contrast ratio of tumours to normal tissue they should be assessed relative to the signal in the normal tissue of the same prostate region.

The results showed a significant increase in both sensitivity and specificity when using ADC maps alone or with DWI images for diagnostic assessment, as opposed to purely DWI. There are many advantages to using ADC maps which might explain this change. ADC maps give a more accurate measure of tissue diffusion and are particularly useful in differentiating areas which have high signal on DWI images due to T₂ shine through and thus mimic tumour, such as post-biopsy haemorrhage, and therefore can improve specificity by reducing the overcalling of these lesions.

The retrospective studies investigated men with previously confirmed prostate cancer and therefore the readers knew there was cancer present in each prostate. This may cause the readers to be more liberal with diagnosing suspicious lesions in borderline cases where there were no other lesions in the gland. This would help explain the significantly higher sensitivity. Using radical prostatectomy as the reference standard allows the assessment of individual tumours within the gland, facilitates assessment of more than one tumour lesion, and is a more accurate method of defining tumour. In comparison, TRUS biopsy is 'blind' and only samples a small area of the prostate, with a false negative rate of over 30% [190]. This would lead to increased false positives on imaging, decreasing the specificity as was seen in the sub-group analysis.

In comparison to the previous chapter's QUADAS 2 assessment (Page 69) the reviewers of the included studies changed how they interpreted the limitations of using patients who have had a radical prostatectomy. In the high b-value DWI meta-analysis the risk of bias was high in the

patient selection parameter. However, in this meta-analysis the high risk of bias shifted to high applicability concerns in the patient selection parameter (Page 88).

To conduct the ideal blinded study there should be assessment of the diagnostic performance of an imaging sequence against a reference standard in a population with and without the disease. This is not possible if the reference standard involves removing the prostate as in patients who have had a radical prostatectomy. Initially, it was believed that this study design limitation introduced a high risk of bias because according to the guidelines on performing the QUADAS-2 assessment there were “inappropriate exclusions” in patient selection. However, on further reflection these studies did not have inappropriate exclusions, but rather the patients “included in the study differ, compared to those targeted by the review question” [122] thus making the patient selection less applicable.

Arguments could be made that using only patients who have had a radical prostatectomy undermines the applicability of the diagnostic results or introduces significant bias into the study, or both. It is perhaps more important to note that many studies used radical prostatectomy specimens and highlight the strengths and weaknesses of this. This is because it allows assessment of the whole gland and for multiple tumours, and it provides a means of more accurately comparing visual assessment of tumour radiologically and histopathologically as both are ‘sliced’ in or close to the axial plane. The fact that all the patients have cancer, tend to be younger, and have higher grade disease, could influence the reading of the MRI by the radiologists and give a falsely high assessment of the diagnostic performance.

This systematic review has a few limitations. Like the previous chapter’s review the search was limited by a finite number of databases, however, those chosen contain a large amount of the relevant journals and by exploring the grey literature and hand searching references the search strategy was of sufficient sensitivity. Secondly the search was limited to the English language, a majority of articles are published in English but there may be data in another language that were

not include in this meta-analysis. Again publication bias was not assessed for reasons stated in the statistical analysis section, however the degree to which publication bias impacts diagnostic tests is unknown [191]. The exact T2WI parameters for the included studies were not assessed which could explain some of the heterogeneity seen. Reader experience is another factor which was not assessed as it was often poorly reported and in different formats such as years practicing, years reporting prostate mpMRI, or number of prostate mpMRIs. It is recognised that reader experience is important in interpreting mpMRI and should be considered when implementing prostate imaging [192,193]. Although diagnostic accuracy is important for the prostate cancer assessment there are other aims of mpMRI which have not been assessed in this meta-analysis. For example, assessment of extracapsular extension, seminal vesicle or lymph node involvement as well as its ability to quantify tumour size and volume. These findings are all used for proper staging of disease and need to be considered when deciding optimal imaging sequences. Finally, accuracy is but one measure of the usefulness of a diagnostic test. A test should also be consistent and therefore interobserver reliability assessments should be performed when studying a novel diagnostic technique.

In conclusion, the advantage of this meta-analysis over the previous is the greater number of included studies. The primary aim of many of the included studies was to specifically assess DWI sequences and techniques, but also included T2WI. This is closer to typical clinical assessment for tumour on prostate MRI, particularly using a visual assessment for tumour. The diagnostic accuracy of diffusion-weighted and T₂-weighted imaging for prostate cancer detection is good. There is improved diagnostic accuracy using b values of 1,000 sec/mm² and greater, but the results of the b>1000 subgroup should be interpreted cautiously given the small number of included studies. Indeed, all the pooling results should be carefully scrutinized as there was a high degree of statistical heterogeneity between the included studies, however, the quality of included studies was considered good by QUADAS-2 assessment.

Summary of meta-analyses and aims of FODIP study

The literature and results of the two meta-analyses performed in Chapters 2 (Page 55) and 3 (Page 82) suggest a role for higher b-values, particularly at 3T, a recommendation which is echoed in the latest PI-RADS version 2 guideline for imaging and reporting prostate MRI [67]. This iteration of the guidelines suggests using a maximum b-value of 1400-2000 sec/mm² if the MRI technology, such as the scanner, software and manufacturer, permits a high enough SNR. The lowest b-value should not be 0 sec/mm² but rather 50–200 sec/mm² to negate the pseudo-perfusional effects as described in the introduction.

The results of the meta-analyses, and review of the literature and current guidelines resulted in the implementation of a non-zero low b-value, namely 100 sec/mm², and the use of two high b-values, 1000 sec/mm² and 1500 sec/mm², for the DWI sequences in the FODIP study.

Small FOV techniques demonstrate promise, however, there is currently no study of the diagnostic accuracy and quantitative diagnostic parameters of small FOV DWI techniques with radical prostatectomy specimen as the reference standard. Furthermore, with the literature and guidelines promoting the use of high b-values, no research group has studied a DWI sequence which combines the use of high b-value and small FOV techniques.

Proceeding with the FODIP study the results of the meta-analyses justified the plan for a prospectively designed observational diagnostic test accuracy study. Patient selection is important and although there is bias in using patients who have had a radical prostatectomy, the advantages of better radiologic-pathologic correlation outweigh these. Additionally, by using experienced readers (radiologists and histopathologists) blinded to clinical details the diagnostic performance of a novel small field-of-view DWI sequence could be assessed, whilst also comparing it to a more conventional larger field-of-view sequence and also the results of these meta-analyses. The

results also rationalise the use of b-values greater than 1000 sec/mm^2 , the assessment of DWI alone and in combination with T2WI, and the use of DWI and ADC maps in combination.

The aim of the FODIP study is to determine the diagnostic accuracy and reliability of a small FOV sequence (FOCUS), performed at the b-values described above, and compare the performance to a conventional DWI sequence with matching b-values.

Chapter 4 Materials and Methods

Summary

This was a prospectively designed observational study of the diagnostic accuracy and reliability of two diffusion-weighted sequences at detecting prostate cancer. The sequences compared were a conventional large FOV EPI DWI sequence in which the whole pelvis is imaged and a smaller FOV EPI DWI sequence in which the imaging field-of-view is confined to the prostate and immediate adjacent tissue. Henceforth the sequences will be known as cDWI (conventional DWI^c) and sDWI (small field-of-view DWI). The cDWI and sDWI sequences had other technical parameters matched, such as b-values used and slice thickness. The MRI were read by two radiologists blinded to clinical, biochemical and histopathological details. The accuracy was determined with respect to radical prostatectomy specimens, which was reviewed by two pathologists, with all lesions measuring 5mm or greater in maximum axial diameter assessed.

Research questions

1. What was the diagnostic accuracy of sDWI and cDWI sequences at detecting and locating prostate cancer alone and in combination with T2WI sequence?
2. What was the inter-observer reliability of the cDWI and sDWI sequences alone and in combination with T2WI sequence?
3. Did the ADC values of tumours differ between the cDWI and sDWI ADC maps? Was there a correlation between ADC value and Gleason Score and Gleason Grade Group?
4. What were the quantitative characteristics of tumours on sDWI and cDWI sequences, namely signal-to-noise ratio (SNR), contrast ratio (CR) and contrast-to-noise ratio (CNR).

^c The use of 'conventional' does not represent a specific type of DWI sequence which is available and widely used, but refers to a more typical EPI DWI sequence. 'Conventional' is used rather than 'large FOV' to avoid the emphasis on the difference between the FOV.

Routine clinical care of patients with prostate cancer at NNUH

Clinical pathway

Patients who are suspected of having prostate cancer are usually assessed with clinical examination, blood tests and prostate biopsy (usually TRUS biopsy). Patients who have a positive biopsy and are being considered for radical treatment are offered a MRI. In addition, patients who are considered at high risk of having prostate cancer, but have had a negative biopsy are recommended to have a MRI to help guide a targeted biopsy.

In the NNUH there are 4 MRI scanners on which MRI prostate studies are performed. The 3T field strength MRI scanner at NNUH has the sDWI sequence. Patients can be scheduled for their prostate MRI scan on any of the scanners. The MRI prostate protocol for all patients being scanned on the 3T MRI is described below. MRI source data, including scan images, are stored on the NNUH NHS picture archiving and communications system (PACS), which is a secure database.

The delay between biopsy and MRI at NNUH is recommended to be 6 weeks [68]. This is to allow the post-biopsy haemorrhage to settle in the gland to allow better assessment of the MRI. The NHS mandates a 62 day timeframe for cancer patient management from initial referral to treatment. Waiting 42 days in the middle of the pathway for haemorrhage within the prostate gland to settle would make the target difficult to achieve. A compromise, agreed between the radiology and urology department at NNUH, is to have a 4 week delay between biopsy and MRI, which in general is adhered to.

After issuing of the MRI report patients are discussed at the weekly uro-oncology meeting at NNUH. Based on the results, clinical assessment, MDT consensus, and patient wishes, some patients will go on to have radical prostatectomy to treat their prostate cancer.

The radical prostatectomy specimens of all patients are analysed with a technique known as whole mount step section. Prostates can either be completely or partially sampled after prostatectomy. Whole mount section is a recognised method of complete prostate sampling [127]. After surgery the prostate is stored overnight in formaldehyde. The seminal vesicles are removed and the prostate sliced from base to apex in the axial plane, with the apical slice then sliced in a longitudinal manner to assess for tumour in the extreme apex. The sections are stained with haematoxylin and eosin and analysed for presence and location of tumour, Gleason score, extracapsular extension and seminal vesicle invasion.

Routine clinical MRI Protocol

MRI prostate studies performed on a 3T MRI (Discovery MR750w wide bore, GE Healthcare Systems, Milwaukee, Wisconsin) at NNUH adhere to the following protocol, which was implemented on 1st September 2015.

- Standard anatomical sequences include axial T1WI and T2WI of the whole pelvis; coronal T2WI fat-saturated sequence of the pelvis; high resolution axial, sagittal and coronal T2WI sequences of the prostate and seminal vesicles.
- Axial DWI sequences include cDWI of the whole pelvis, and sDWI (known as FOCUS by manufacturer) sequence of the prostate and seminal vesicles.
- A quadrature body coil for transmission and a 36-channel phased-array receiver coil. No endorectal coil is used.

The order of sequences described above is the order the sequences are performed. 20mg of intramuscular hyoscine butylbromide (Buscopan, Boehringer Ingelheim Ltd., Bracknell, United Kingdom) is administered following the coronal T2WI fat-saturated sequence.

Total scan time is approximately 40 minutes.

Study procedure

Recruitment policy

Recruitment continued until 40 eligible patients completed the study.

Inclusion criteria

- Males
- Aged 18 and over
- Biopsy proven prostate cancer.
- No contraindication to MRI scanning.
- Smokers and non-smokers
- Patients who had a complete scan on the 3T scan MRI at NNUH and then went on to have a radical prostatectomy at the NNUH.

Exclusion criteria

- Those regularly taking 5-alpha reductase inhibitors or testosterone replacement medicines.
- Those with significant artefact from surgically implanted metalwork.
- Those who have had pelvic radiotherapy.
- Those with a delay of less than 21 days between biopsy and MRI.

Identification of patients

There was a retrospective review of patients who have had a MRI at 3T at NNUH from 1st September 2015 on Soliton, the Radiology Information System.

There was a retrospective review of patients who had a radical prostatectomy at NNUH from 1st September 2015 using the trust theatre booking system Orsos.

Patients who had a MRI prostate examination at 3T and had a radical prostatectomy will be considered for the study. Following this there was a review of patient clinic notes and previous imaging to determine if they met the other eligibility requirements.

Patient characteristics

The patient characteristics data collected included name, patient number, date of birth, age at MRI, date of MRI, sequences performed, missing sequences, indication for MRI, PSA result (prior to biopsy if possible), biopsy performed, date of biopsy, biopsy type, biopsy positive, time from biopsy to MRI, Gleason score from biopsy, side of gland with tumour, date of radical prostatectomy, and time from MRI to radical prostatectomy.

MRI imaging parameters

Once the required number of eligible patients was reached the MRI data was collected and analysed. There was no interference with the routine clinical pathway of the patients eligible for the study.

The sequences relevant to the study of eligible patients' MRI studies were placed in a unique anonymised folder on the local PACS system, a secure database, with only each patients' unique study number displayed (described in the section 'Patient Confidentiality'). No other identifiable information was displayed. This was important ethically as the radiologists who read the MRI for the study were not necessarily part of the clinical team involved in the patients' care when the studies were initially read prior to the MDT and therefore would not be allowed access to their images and personal data. In addition, if the radiologist was involved in the initial read of the MRI as part of patients' routine clinical care, anonymising patient data and using a study number reduced the risk of recall bias.

The parameters of the sequences relevant to the study are displaced in Table 4-1.

Table 4-1. Imaging parameters of sequences relevant to the FODIP study.

Parameter	Sequence			
	T1WI	HiRes T2 PROPELLAR (T2WI)	Conventional DWI (cDWI)	FOCUS DWI (sDWI)
TR (ms)	450	9307.14	7000	4000
TE (ms)	12.86	107.07	77	81
Bandwidth (Hz)	31.25	83.3	250	250
Echo-train length	3	26	N/A	N/A
Matrix size	320P/512F	320/320	80/92	160/80
FOV (cm)	34 x 40.7 (38cm)	24 x 28.7 (24cm)	34x40.7 (34cm)	26 x 13 (13cm)
No. of acquisitions	1	2	1,2,3 (per b-value)	8,12,16 (per b-value)
Slice thickness (mm)	6	3.60	3.60	3.60
Interslice gap (mm)	2	0	0	0
Parallel imaging factor (ARC)	1.77	2	2	N/A
b-value (sec/mm ²)	N/A	N/A	100, 1000, 1500	100, 1000, 1500
ADC maps	N/A	N/A	1500, 1000, 100 (confidence level 0.4)	1500, 1000, 100 (confidence level 0.4)
Duration (min.sec)	4.00	5.30	6.30	8.00

TE – Echo time, TR – Time of repetition

MRI assessment

Two radiologists, Dr Tristan Barrett (Rad1/Reader 1) and Dr Paul Malcolm (Rad2/Reader 2), assessed the MRI studies. Both are experienced uro-radiologists with 7 and 15 years of experience reading prostate MRI, respectively. The reads of the MRI were performed independently. They were blinded to clinical and biochemical information, and MRI and pathology findings. They were aware that the patients had prostate cancer and were treated with a radical prostatectomy.

Only the relevant sequences for each imaging protocol were available to the reader. The only patient identifier was a unique patient code. Each patient had a different unique code displayed on each of the MRI assessments. This code only had meaning to KG and not the readers.

There were 4 different imaging protocols:

Protocol A	T1WI, cDWI (inc. ADC map)
Protocol B	T1WI, T2WI, cDWI (inc. ADC map)
Protocol C	T1WI, sDWI (inc. ADC map)
Protocol D	T1WI, T2WI, sDWI (inc. ADC map)

For protocols A and B the cases were read in chronological order of date of radical prostatectomy. For the second read (protocols C and D) the cases were randomised using an online randomiser [194], therefore the cDWI and sDWI cases were read in a different order. By randomising the order the risk of recall and learning bias during the second read was reduced. Protocols B and D were read at the same sitting as A and C, respectively, on a case by case basis rather than after reading all 40 cases. For example, after reading case 32A the reader assessed 32B before moving to 33A.

For each protocol read the T1WI sequences were available to assess for post-biopsy haemorrhage. Haemorrhage was defined as an area of high signal on the T1WI sequence. Presence and location of haemorrhage was documented on the reporting template. Figure 4-1 illustrates the high signal caused by haemorrhage, usually following TRUS biopsy, within the peripheral zone.

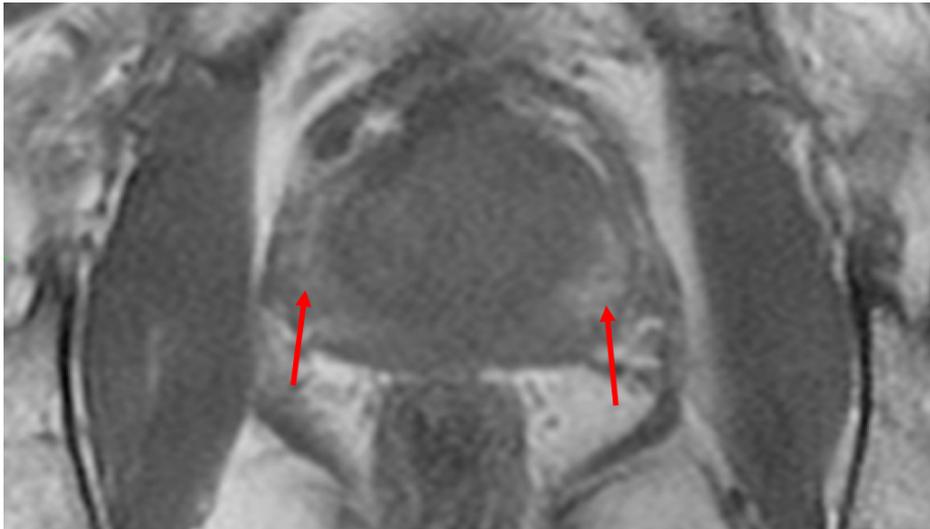


Figure 4-1. T1 weighted image of the prostate demonstrating post-biopsy haemorrhage in the peripheral zone bilaterally (red arrows).

There was a delay of at least 2 weeks between the read of the cDWI protocols (A and B) and the sDWI protocols (C and D) to reduce recall bias. The MRI and histopathology assessment timeline is summarised in Figure 4-2.

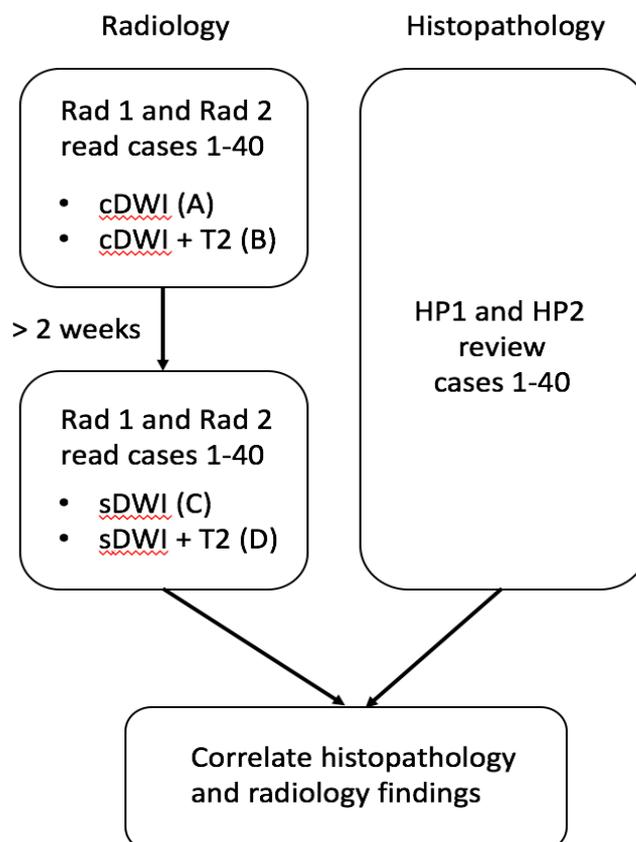


Figure 4-2. Histopathology and imaging assessment timeline.

On T2WI sequences tumours have different appearances depending on whether they are located in the peripheral or transitional zone (PZ and TZ) [67]. Within the TZ a homogenous low signal mass which destroys normal structures or benign pathology such as hyperplastic nodules in benign prostatic hyperplasia, is considered definitely malignant. Within the PZ, lesions which exert mass effect, are homogeneously low signal, and are irregular in shape with ill-defined borders, are considered definitely malignant. An example of prostate cancer in the PZ and TZ on T2WI is displayed in Figure 4-3.

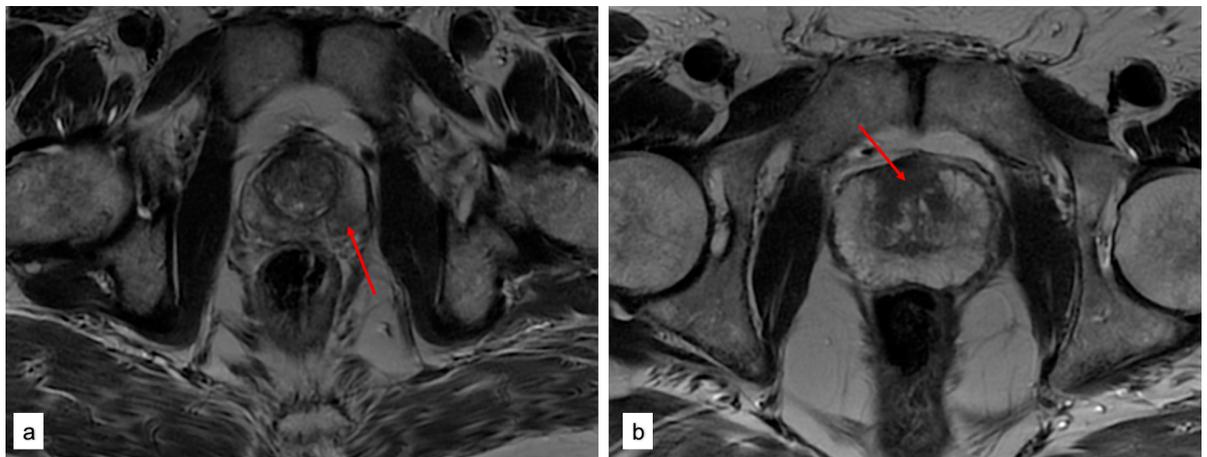


Figure 4-3. Example of malignant lesions on T2WI axial images. a) A focus of Gleason Score 3+4=7 prostate cancer in the left peripheral zone at 5 o'clock. b) A focus of Gleason Score 4+3=7 prostate cancer in the transitional zone anteriorly (red arrow).

On DWI, a lesion demonstrating high signal on the b-value imaging with corresponding low signal on the ADC map is considered malignant. An example of prostate cancer on the different DWI images and ADC maps is demonstrated in Figure 4-4.

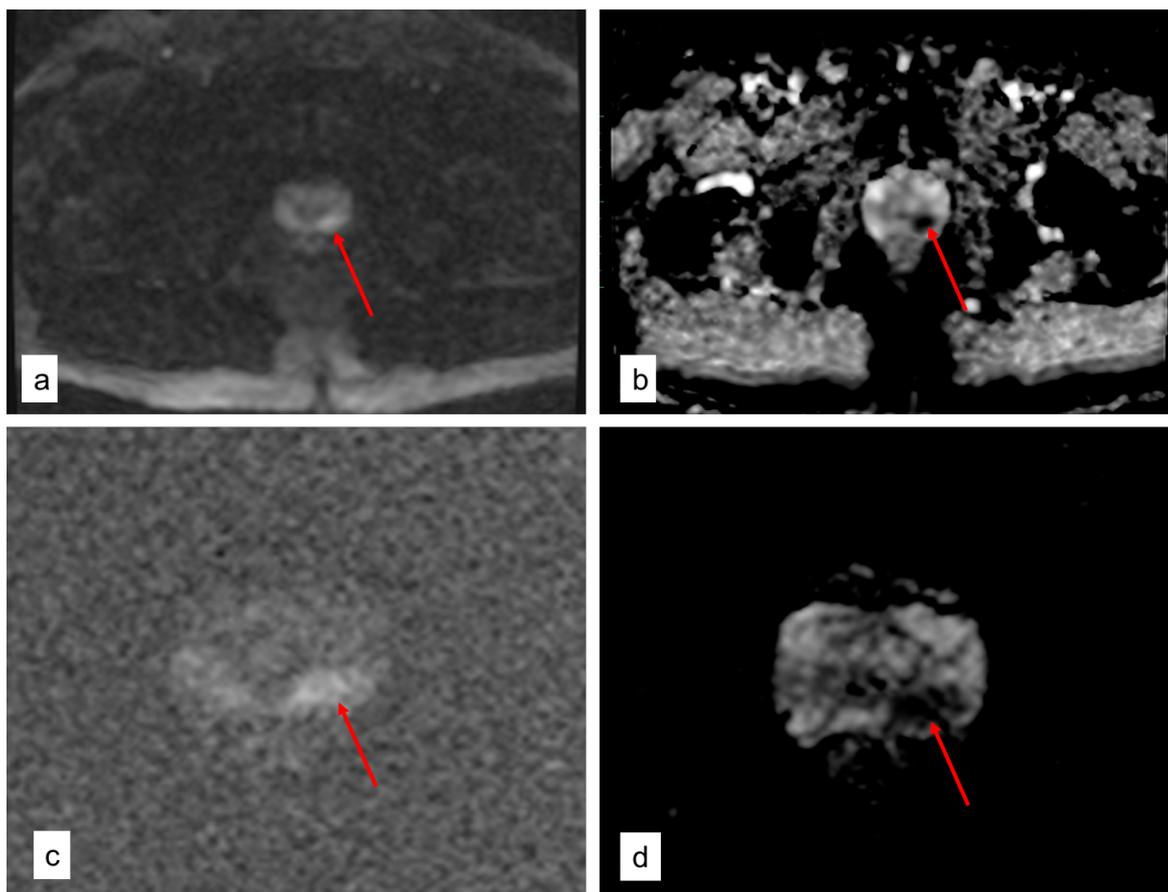


Figure 4-4. Example of the same left peripheral zone prostate tumour on different DWI sequences and their respective ADC maps. On the DWI images the tumour appears as high signal relative to surrounding non-tumour tissue and on ADC maps the tumour is lower signal than non-tumour tissue. The same Gleason Score 4+3=7 tumour is demonstrated on a) b1500 cDWI. b)ADC map cDWI. c) b1500 sDWI. d) ADC map sDWI.

Radiologist assessment

The lesion and sectoral-based accuracy, as described in the introduction to chapter 2 (page 56), were assessed for both readers and all protocols.

The radiologists documented the location of each tumour in terms of right or left hemi-gland; the base, mid-gland or apex; and the TZ and PZ. The maximum axial diameter and the width perpendicular to the maximum axial diameter on the same image slice were documented. Finally, the top and bottom slice the tumour was seen on was noted. A Likert score for the lesion was also documented as well as the presence of extension beyond the capsule or seminal vesicle invasion. The lesion was also drawn on the prostate reporting diagram Figure 4-5 taking care to demonstrate the sectors the lesion involves.

The lesions identified by the radiologists were assessed by a 'stringent' and 'clinical' approach. A 'stringent' assessment required the reader to correctly identify the tumour in the exact location it was identified on histopathology. A 'clinical' approach considered a lesion to be a true positive if it was in a neighbouring sector as long as the correct hemi-gland was selected. The 'clinical' approach allowed for some leeway, as although a radiologist could have correctly identified the tumour the histopathologists determined the tumour to be in a different location, possibly due to the limitations in the correlation of MRI and histopathology. Therefore, following this approach the localisation of the tumour by the reader was assessed as being 'close enough' thus putting more emphasis on the detection of the tumour rather than localisation.

The sectoral accuracy was determined by assessing the tumour reporting diagram (Figure 4-5). A 12 sector reporting diagram was used with the gland divided into left and right; apex, mid-gland and base; and peripheral and transitional zones. For each sector, each reader provided a score of 1-5 depending on the likelihood of tumour in this region. The Likert scores used in the lesion and sectoral assessment were as follows:

1 - definitely absent

2 - probably absent

3 - indeterminate

4 - probably present

5 - definitely present

The reporting instructions for the radiologists are in Appendix 2 (Page 262).

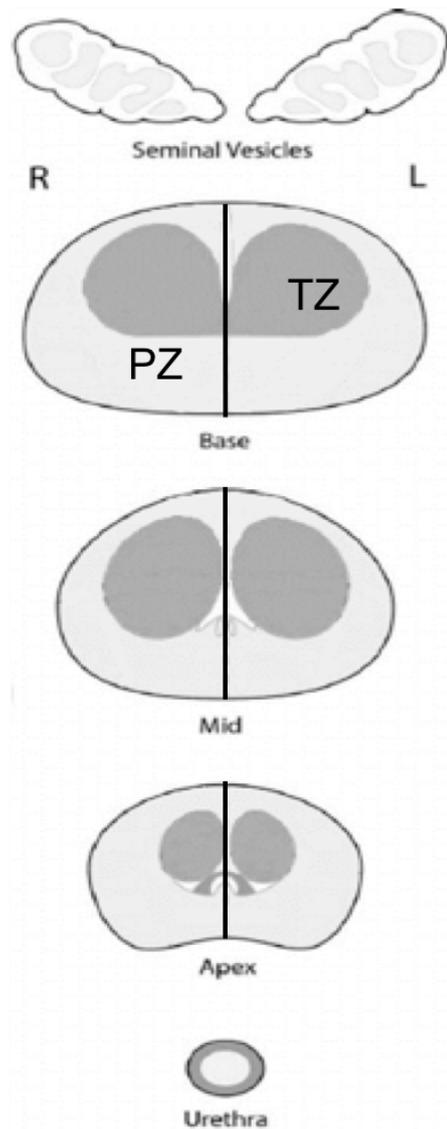
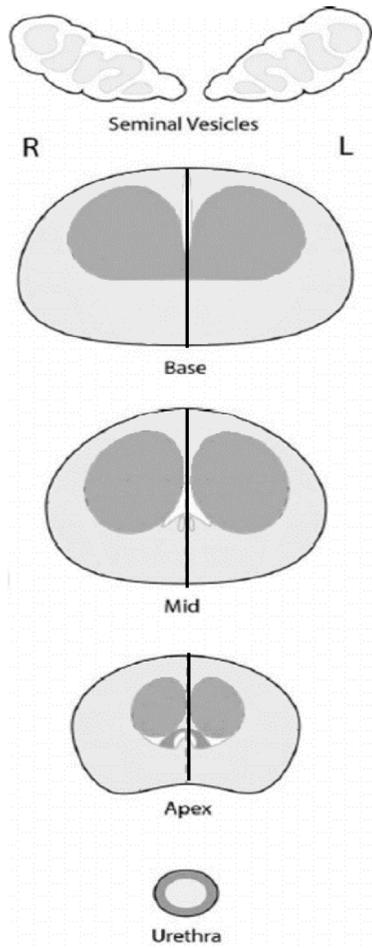


Figure 4-5. Prostate reporting diagram divided into 12 sections.

Radiologist reporting template

The reporting template is displayed in Figure 4-6. An example of a completed radiologist reporting template is demonstrated in Figure 4-7. The completed template is by Rad2 (PNM) demonstrating a lesion in the anterior gland which was interpreted to be highly suspicious for prostate cancer. The reader thought the tumour extended from apex to base and resided in both sides of the gland and in the peripheral and transitional zones. Each of the 12 sectors was scored a 5, as denoted by the Likert score column. The striped area indicates there is haemorrhage throughout the right peripheral zone.



Case Number:

Reader:

Haemorrhage: Y / N

Lesion	Location			Max Axial Diameter (mm)	Series/ Slice No.	Width (mm)	Height (slice)		Likert (1-5)	ECE (Y/N) (mm)	SVI (Y/N)
	B/M/A	R / L	PZ/TZ				Top	Bottom			
1											
2											
3											
4											
5											

Figure 4-6. Radiology reporting template.

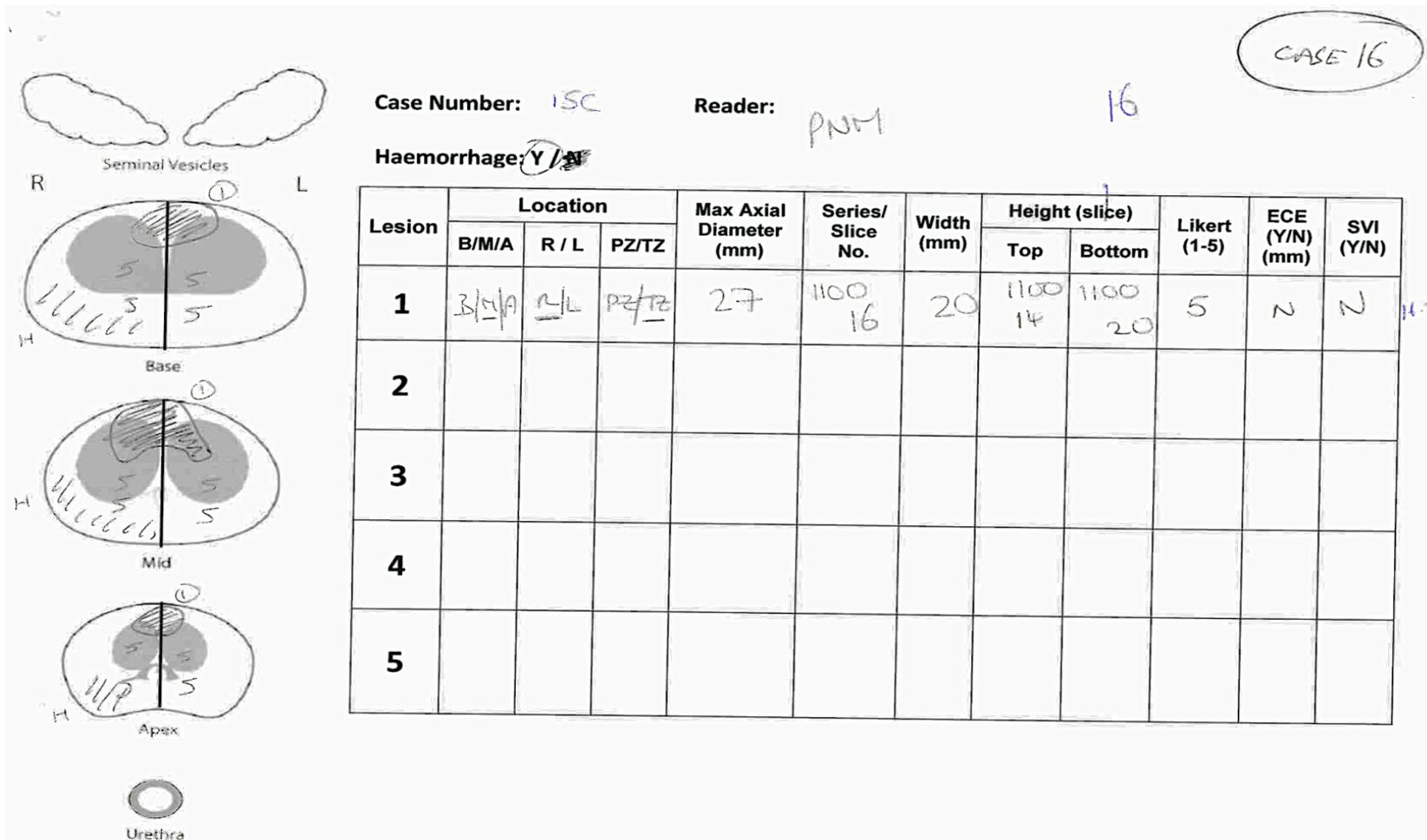


Figure 4-7. Completed radiologist template for case 16.

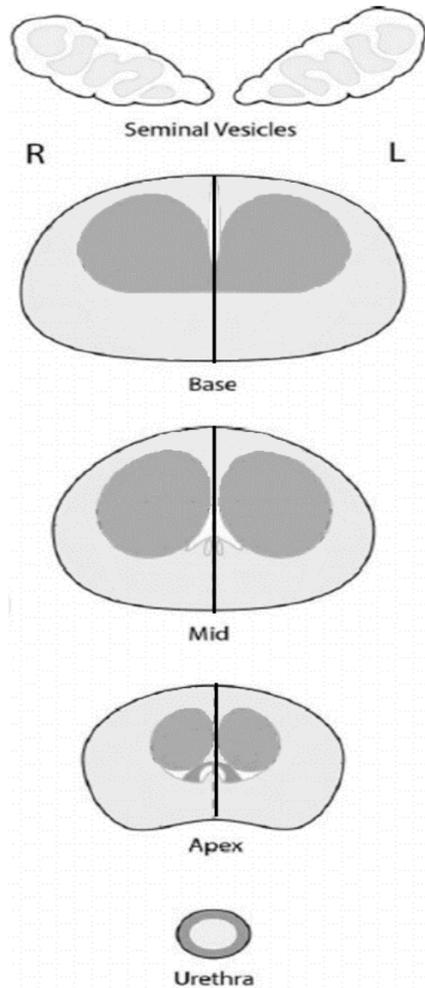
Histopathology analysis

As mentioned above (in the 'Standard clinical care' section) all patients who have undergone radical prostatectomy had their prostatectomy specimens evaluated with a technique known as whole mount step section.

The radical prostatectomy specimens of eligible patients were re-assessed for this study. Two histopathologists (HP1 and HP2) performed this independently. HP1 is an experienced consultant uropathologist with 20 years experience, and HP2 is a senior histopathology trainee with a particular interest in uropathology. The histopathologists were blinded to the clinical, biochemistry and MRI findings. It was not possible to blind the histopathologists to patient name, unique hospital number and date of birth. These details are written on sample pots and slides making blinding of this information difficult to achieve. Furthermore, blinding could introduce a significant risk of error in sample identification and mis-sampling.

As for the routine clinical histopathology assessment the Gleason score, presence of extracapsular extension and seminal vesicle invasion of each lesion were assessed, however, lesions less than 5mm in maximum axial diameter were excluded. The location of the tumour with reference to the 12 sectors was documented.

The histopathology assessment will be reported on a histopathology reporting template (Figure 4-8) with the patients' unique study number as the only identifying patient detail. The similarity of the radiology and histopathology template will facilitate correlation of the two reports.



Study Number _____

Reader or consensus _____

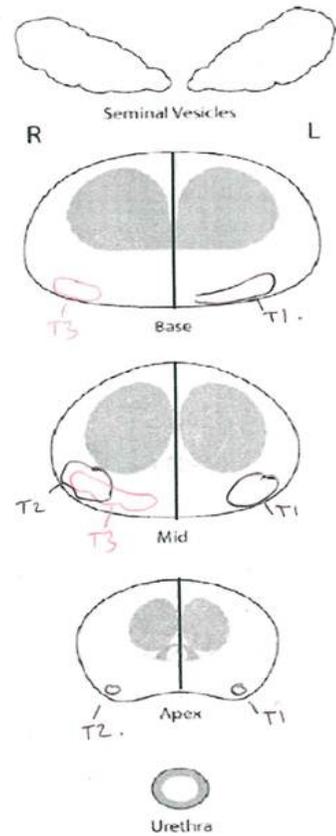
Lesion	Lesion location			Maximum width diameter of prostate =			Estimated slice thickness =		
	B/M/A	R/L	PZ/TZ	Maximum axial diameter (mm)	Width (mm)	Height (mm)	ECE (mm) Y/N	SVI Y/N	Gleason/ Grade Grp
1									
2									
3									
4									
5									
6									

Figure 4-8. Histopathologist reporting template.

No extra tests (stains, genotyping etc) were performed on the specimen and did not prevent further clinical examination of the tissue in the future.

The definitions of the dimensions were: the length was the maximum dimension of the tumour on the prostate slide where the tumour was visually the longest; the width was measured perpendicular to the length on the same slide; and the height was the estimated height of the prostate based on the number of slides the tumour was visualized on considering the estimated distance between slides. The estimated slice thickness was calculated by dividing the height of the prostate by the number of slices. The height of the gland is documented as part of the initial assessment of the prostate following prostatectomy within standard NHS care. The reporting instructions for the histopathologists can be found in Appendix 2 (page 268).

An example of a histopathological assessment of a prostatectomy specimen with a completed histopathology template (Figure 4-9) and the photograph of the pathology specimens with the tumours outlined (Figure 4-10).



Study Number 3

Reader or consensus Consensus

Lesion	Lesion location			Maximum width diameter of prostate =			Estimated slice thickness = 6.5		
	B/M/A	R/L	PZ/TZ	Maximum axial diameter (mm)	Width (mm)	Height (mm)	ECE (mm) Y/N	SVI Y/N	Gleason/Grade Grp
1	BMA	L	PZ	12	8	27	N	N	3+4=7/2
				11	8	35			
2	MA	R	PZ	14	11	11	N	N	3+4=7/2
				14	10	13			
3	MIB	R	PZ	23	8	20	N	N	3+4=7/2
				22	8	32			
4									
5									
6									

T2 Stops in caudal half of ~~middle~~ middle.
 T2 starts in caudal half of middle

Figure 4-9. Completed histopathologist template for case 3.

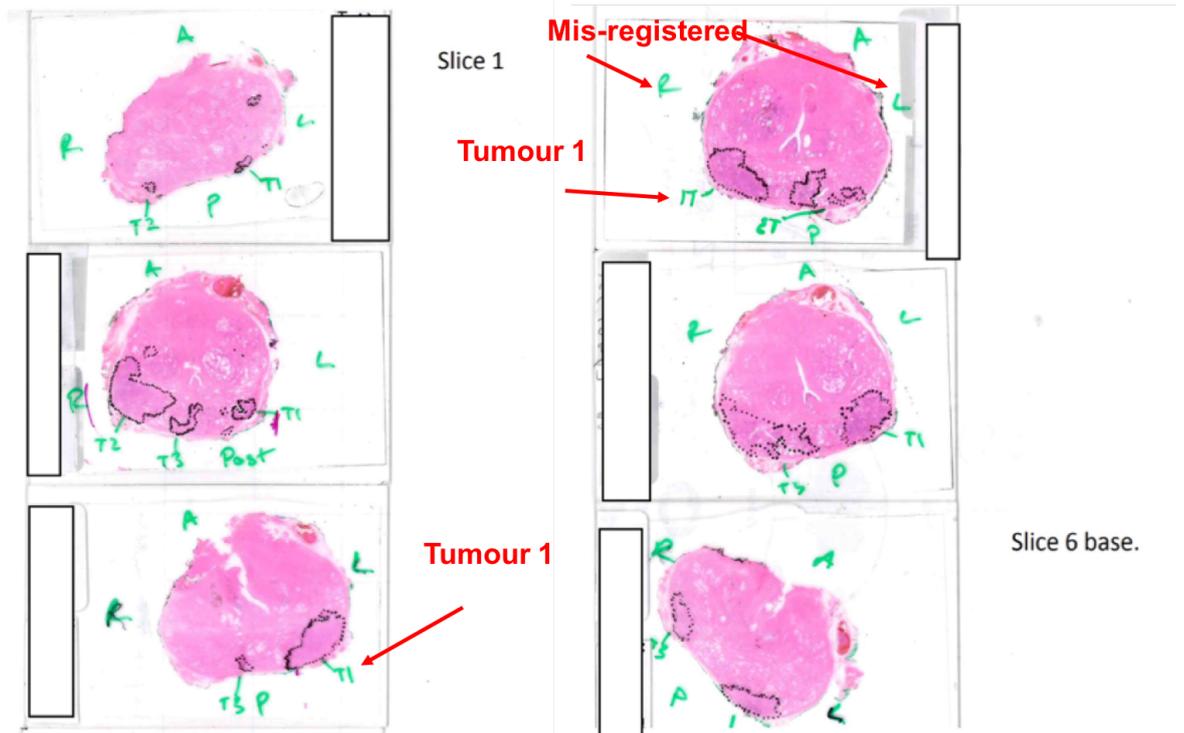


Figure 4-10. Case 3 histopathology slides.

Figure 4-9 and Figure 4-10 demonstrate the histopathology assessment and results for their analysis of case 3 radical prostatectomy specimen. There were six slices and three eligible tumours. Lesion 1, denoted with the T1 label (highlighted by the red “Tumour 1” marker and arrow), resides in the left peripheral zone at 5 o’clock and is present in apex, mid-gland and base. Lesions 2 and 3 (labelled by the histopathologists as T1WI and T2WI) reside in the right peripheral zone. This example case also highlights the potential for mis-registration as on the 4th slice (top right in Figure 4-10) the right (R) and left (L) labels were drawn incorrectly (highlighted by the red “Mis-registration” marker and arrow). Following discussion between KG and HP2 it was decided that the left and right were labelled incorrectly.

Once the samples have been analysed they were returned to storage, as they would have done following the initial routine clinical assessment following prostatectomy.

Correlation of MRI and Histopathology

1. Initial review of pathology results against images

After the histopathology templates were completed KG reviewed the sectoral maps, drawings of tumours and tumour parameters outlined in the histopathology reported template of all cases, and compared these against the MRI images. This initial step prior to determining accuracy aimed to, firstly, assess any cases with multiple tumours and determine on imaging whether multiple tumours could all be part of one tumour. The advantage of the DWI and T2WI sequences was the absence of gaps between the images (3.6mm slice thickness and no gap) ensuring the entire prostate had been imaged. On the other hand, the pathology assessment is of a micrometre thin slice every 4 to 6mm, making the potential for a tumour to extend or stop between slices a possibility. Therefore, a case in which there were thought to have been multiple lesions on histopathology may actually have been a single lesion as they 'joined up' on imaging. Secondly, all lesions will be assessed to determine whether they meet eligibility requirements, most importantly if they meet size criteria.

2. Histopathology and radiology consensus of contentious cases and missed lesions.

Thereafter, KG and HP2, in consensus, assessed those contentious cases outlined in the step above with the images, pathology slides, and histopathology templates available. If a multifocal tumour is decided to represent one lesion then the pathologists amended the tumour dimensions and prostate map appropriately to reflect this.

Another important function of the review of images and histopathology results was to decide the location of tumours on imaging that were present on pathology and eligible for inclusion but cannot be identified on cDWI and/or sDWI. It was important to perform the qualitative and quantitative analysis of all lesions, including the misses, to attempt to determine why these

lesions were not seen. Reasons for radiologist missing lesions might include the presence of haemorrhage, a small tumour and a low Gleason grade. The histopathology data was then entered into a spreadsheet to be analysed.

3. Sectoral accuracy of MRI interpretation

From the histopathology tumour maps it was possible to determine which of the sectors of prostate had tumour present. Each sector was given a binary code to represent whether tumour was present or not.

This was then compared (by KG) against each radiologist's Likert score for each sector for each of the four protocols. From this the 2 x 2 contingency table could be completed for different Likert score cut-offs. The definitions of each component of the 2 x 2 contingency table for each sector were as follows.

True positive (TP)	Radiologist score was higher than the Likert cut-off and the pathologists deemed tumour was present
True negative (TN)	Radiologist score was less than the cut-off and the pathologists deemed tumour was absent
False positive (FP)	Radiologist score was higher than the cut-off and the pathologists deemed tumour was absent
False negative (FN)	Radiologist score was lower than the cut-off and the pathologist deemed tumour was present

From this the sensitivity, specificity, and positive and negative predictive values were calculated using the following formulae. Additionally, the receiver operating characteristic curves (ROC) were created and the area under the curve (AUC) calculated.

$$\textit{Sensitivity} = TP / (TP + FN)$$

$$\textit{Specificity} = TN / (TN + FP)$$

$$\textit{Positive predictive value} = TP / (TP + FP)$$

$$\textit{Negative predictive value} = TN / (TN + FN)$$

4. *Lesion accuracy of MRI interpretation*

The lesion accuracy analysis of the histopathology and radiology reports was performed by KG. From assessment of the histopathology tumour maps and the radiology tumour maps it was possible to determine three components of the 2 x 2 contingency table: true positive, false positive, and false negative. Following assessment of the diagnostic accuracy results of the sectoral analysis by KG a Likert cut-off of 3 and above for a tumour was considered to represent a possible tumour and those with a score of 2 or less disregarded. A justification for this cut-off is discussed in Chapter 6 (page 194). A 'stringent' and 'clinical' lesion accuracy analysis was performed.

A true positive lesion for the 'stringent' approach was a radiologist determined lesion with a Likert score of 3 or above which was present in the same sector as a lesion on the pathology tumour map. In addition, the lesion visually must overlap on the radiologist and pathologist prostate map and this is at the discretion of KG. This prevented tumours at different extremes of a sector being considered a true positive, for example a 7 o'clock PZ lesion on the radiological map and a 11 o'clock PZ lesion on the pathological map being considered a correctly identified tumour rather than a false positive lesion.

The 'clinical' approach allowed tumours which overlap in appearance, but are up to one sector out in height or PZ or TZ, but confined to the correct hemi-gland to be considered a true positive.

A false negative for both approaches is a tumour seen on the pathology map but not seen by the radiologist.

A false positive is a tumour described by a radiologist but not seen by a pathologist.

Using the results the sensitivity and positive predictive values of each radiologist for each protocol was calculated using the equations demonstrated above.

5. Lesion accuracy of index lesions

The index lesion was defined as per the ISUP guidelines [43] based on the following priority: presence of extension beyond the prostate, then tumour grade, then tumour volume. For each radiologist and on each protocol whether the index lesion was correctly identified (true positive) was determined by KG using the 'stringent' and 'clinical' definition outlined above.

6. Quantitative analysis: Contrast-to-noise ratio , Contrast ratio and Signal-to-noise ratio

The anonymised cases were transferred to a PACS known as Horos (version 1.1.7). This PACS software allowed a closed polygonal shape to be drawn over any part of the image and also any slice. The closed polygonal shape is referred to as a Region-of-Interest (ROI). It also allowed splitting of the DWI images into their unique b-values, namely b100, b1000 and b1500. These individual b-value images, ADC maps, and T2WI and T1WI images are linked anatomically so scrolling on one set of images results in scrolling through the images of the others automatically.

ROIs drawn on one image can be copied to the exact location on all other sequences. In some instances, there were some mis-registration between sequences resulting in the copied ROI not residing in the exact anatomical location as the initial ROI. When this occurred a judgement was made by KG whether they matched and moved to a more accurate position if required.

Horos PACS ROI provided multiple data parameters including the mean, maximum, minimum and standard deviation of the signal intensity within the ROI. The mean signal intensity on the ADC map correlates to the mean ADC of the ROI. In addition, the surface area of the ROI is provided. The signal intensity is denoted by the symbol μ . The images of every ROI for all cases were saved.

Figure 4-11 demonstrates case 19 lesion 1 on sDWI images and T2WI and T1WI axial sequences. Following assessment of the histopathology tumour map a ROI was drawn around, but not outside, the margin of the tumour taking care not to extend into non-tumour tissue. The ROI was initially drawn on b1500 image as this was the sequence that the tumour was most conspicuous and the ROIs were copied to all other b-value images and the ADC map. Data from each ROI provided by the imaging software was collected. Haemorrhage was present in the peripheral zone as indicated by the high signal on the T1WI sequence (bottom right image), but was not present at the location of tumour.

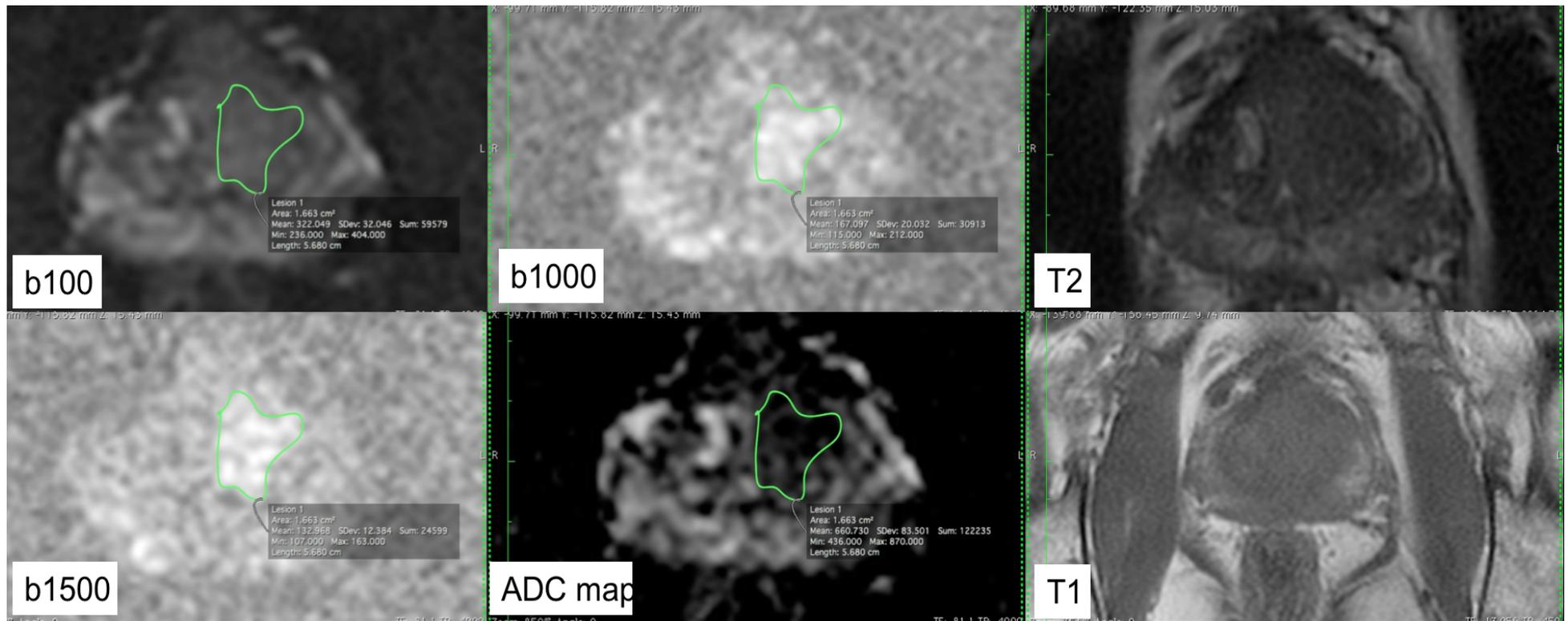


Figure 4-11. Left transitional zone prostate cancer with ROI drawn around the lesion on all three b-value images and ADC map. The corresponding T2WI (top right) and T1WI weighted slice are also demonstrated.

ROIs were drawn in an area of normal prostate peripheral zone and normal prostate transitional zone. The location of the normal tissue was decided by KG after scrutinizing the histopathology tumour maps and imaging sequences taking care to avoid areas of haemorrhage noted on the T1WI sequence as high signal.

Tumours were identified by KG by consulting the histopathology tumour maps and visualising the images. After reviewing the T1WI, if a tumour was present within an area of haemorrhage this was recorded. When a histopathologically confirmed tumour was visualised on MRI a ROI was drawn around the lesion as seen on the MRI sequence, as opposed to the boundaries defined by the histopathology tumour maps and slides. The first sequence the ROIs were drawn was the sequence the tumour was most conspicuous to KG. The ROI was then copied to other sequences.

For tumours which were not identified on MRI by KG the histopathology map and pathology slides were consulted to best estimate the location. Using local landmarks, such as the urethra, prostate capsule, verumontanum, cysts, calcification and hyperplastic nodules, the ROIs could be more accurately sited [98].

Noise within an image is defined as the standard deviation of the signal within a homogeneous tissue within an image. Many studies estimating noise on prostate MRI are calculated by drawing an ROI within bone [87] or air outside the body [88]. The reduced field-of-view in the sDWI sequence can result in air outwith the patient almost always not visible. Another homogenous tissue in the region is muscle and the closest and most reliably imaged is obturator internus [195]. A ROI was drawn on either the left or right obturator internus.

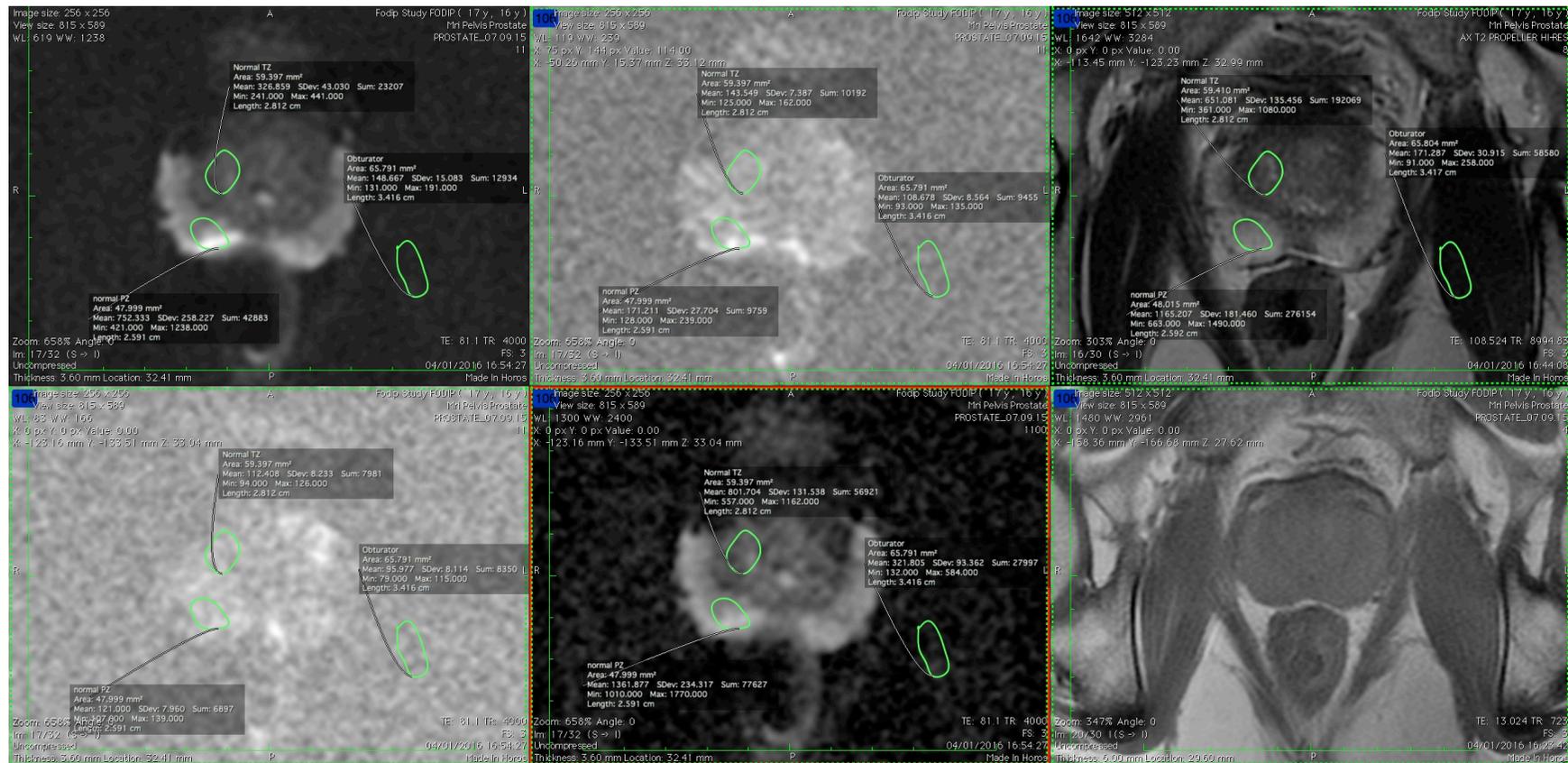


Figure 4-12. sDWI, T2WI and T1WI sequences of case 7. ROI drawn on normal peripheral and transitional zones, and on the left obturator internus muscle.

Figure 4-12 demonstrates ROIs which were drawn on normal peripheral and transitional zones and obturator internus muscle. After consultation of the histopathology tumour map to determine where tumour was not present and assessment of the T1WI sequence to ensure there was no haemorrhage, a peripheral and transitional zone target was identified. The delineation between the PZ and TZ was best appreciated on T2WI and therefore ROIs were copied from the T2WI sequence to ensure they do not extend into the adjacent zone. Once correctly sited they were copied from once sequence to the next and data documented. The obturator internus ROI would initially be drawn on the T2WI sequence and then copied across.

Occasionally, as on the tumour ROI drawing process described above, there was image distortion between sequences resulting in a ROI drawn on T2WI being copied to a different anatomical location on the other sequences. On these occasions the ROI was moved, preserving the shape to a more suitable and accurate location. More frequently peripheral and transitional zone, and obturator internus ROI's were drawn on different slices rather than on the same slice as in the Figure 4-12.

Whether the ROI is being drawn on normal tissue, tumour tissue or muscle, the surface area of the ROI was made as large as possible without extending beyond the boundaries of the intended tissue.

With the signal intensities and noise calculations from the ROI's the signal-to-noise ratio (SNR), contrast ratio (CR) and contrast-to-noise (CNR) were calculated for each tumour by using the equations below. The CNR and CR of tumours was calculated by comparing the contrast between tumour and normal tissue within the zone the tumour predominately resided within. CR is another measure of tumour contrast which does not account for the effect of the noise and is used widely in prostate DWI studies [89,139]. The SNR, CNR and CR were calculated for b1000, b1500 and ADC map images using the following equations.

$$SNR = \frac{\mu \text{ tumour signal}}{\text{standard deviation of noise}}$$

$$CNR(PZ) = \frac{\mu(PZ) \text{ tumour signal} - \mu(PZ) \text{ tissue signal}}{\text{standard deviation of noise}}$$

$$CNR(TZ) = \frac{\mu(TZ) \text{ tumour signal} - \mu(TZ) \text{ tissue signal}}{\text{standard deviation of noise}}$$

$$CR(PZ) = \frac{\mu(PZ) \text{ tumour signal} - \mu(PZ) \text{ tissue signal}}{\mu(PZ) \text{ tumour signal} + \mu(PZ) \text{ tissue signal}}$$

$$CR(TZ) = \frac{\mu(TZ) \text{ tumour signal} - \mu(TZ) \text{ tissue signal}}{\mu(TZ) \text{ tumour signal} + \mu(TZ) \text{ tissue signal}}$$

5. Quantitative analysis – ADC values of tumour

As mentioned previously the mean signal intensity within a ROI on the ADC map also represents the mean ADC of the ROI in mm²/sec. As above, the SNR, CNR and CR of lesions will be calculated for the ADC map to provide a measure of lesion conspicuity on the ADC map. The mean and minimum ADC value of tumours and normal tissue were recorded. Most studies investigating ADC values of prostate cancer use the mean ADC to compare different DWI techniques and measure correlation with Gleason Grade. More recently the minimum ADC value within a ROI has been proven to also have a correlation with Gleason Grade [119,196].

In addition, the mean and minimum ADC values of tumour and normal tissue can be used to create a ROC curve. As in many of the studies included in the high b-value meta-analysis (Chapter 2), ADC values can be used to create a ROC curve and determine the AUC and ideal ADC value cut-off for determining tumour from normal prostate.

Statistical Analysis

Sample size calculation

Using the ADC of tumour of b-0,1000 and b-0,2000 sec/mm² ADC maps in Koo's study [91] analysing DWI in prostate cancer with radical prostatectomy specimens as reference standard at 3T, a calculation was performed to provide a sample representative of the population. Based on a power of 0.90, a significance criterion of 0.05, and a total confidence interval for the mean ADC of 0.1, using the equation outlined by Eng [197], this produced as sample size of n=35. Thirty five patients was rounded up to forty (n=40).

Data analysis

1. Diagnostic performance of small FOV DWI and conventional DWI

The sensitivity, specificity, accuracy, and positive and negative predictive values of each readers assessment of protocols A-D were calculated.

Receiver operating characteristics (ROC) curves were created for each protocol and AUC will be assessed for each radiologist and protocol.

A comparison of the equivalent cDWI and sDWI protocol for each reader was performed using the McNemar test for sensitivity and specificity with Bonferroni correction applied for multiple comparisons, and by the Z-test of proportions for positive and negative predictive values. The comparison of the AUC between cDWI and sDWI was performed using DeLong's test.

2. Comparison of tumour ADC values on different b-values in the same DWI sequence and between sDWI and cDWI datasets.

A Wilcoxon signed rank and paired t-test were used to compare the ADC values of different b-values and between different DWI sequences. ADC differences was assessed for all tumours,

visible and non-visible lesion subgroups, and lesions within and free of haemorrhage subgroups, and this was performed using a Mann-Whitney U test.

3. Compare the SNR, CR and CNR of tumours on different b-values in the same sequence and between sDWI and cDWI datasets.

A Wilcoxon signed rank and a paired t-test were used to compare the SNR, CR and CNR of tumours of different b-values and between the DWI sequences.

4. Comparison of tumour ADC and non-tumour tissue ADC.

A Wilcoxon signed-rank and paired t-test were used to compare ADC and signal intensity on b1000 and b1500 sequences of malignant lesions and non tumour tissue.

5. Accuracy of ADC measurement

Receiver operating characteristic curves were created to determine the AUC and optimal ADC cut-off value between tumour and non-tumour tissue.

6. Correlation between tumour ADC and their Gleason score

The mean and minimum tumour ADC value on cDWI and sDWI was correlated with the Gleason score of the tumour using Spearman Rank Correlation Coefficient test.

7. Inter-observer MRI assessment agreement

Inter-observer agreement of the Likert scores of the sectors between the two radiologists was performed using a weighted kappa statistic. Inter-observer reliability for the peripheral and transitional zones was determined.

Reliability statistics, McNemar, Z-test of proportions and DeLong's test will be done on R statistics (Rstudio version 1.0.143, psych and pROC packages [198]). All other statistics will be performed using SPSS version 23. A p-value of less than 0.05 is considered significant.

Patient Confidentiality

Patient data was collected and stored in a password protected file held in the NNUH IT system, a secure system. It was only made accessible to KG and TSy.

MRI images were stored on the NNUH PACS under the patient name. It was transferred to the BiImaging server sited in the Department of Nuclear Medicine, NNUH. These were accessed from the Image Analysis Laboratory in the NNUH where they were fully anonymised by stripping out all patient identifiable data from the DICOM file header and replacing these data with unique study identifiers. This was conducted by KG. The radiologists reviewing the sequences were blinded to patient identifiers.

Histopathology and radiology reporting forms were kept in a locked cabinet in the radiology department at NNUH and digital copies in a password protected directory in the Trusts IT system.

Ethical Considerations

Consent

Following consultation with Dr Ray Lonsdale, consultant pathologist with a specialist interest in ethics, regarding consent, it was decided written consent was not required for this study. This was for the following reasons

- The routine clinical care of the patient was not altered or interfered with. Although prospectively designed, the analysis of patients' images and pathological specimens

occurred after they had been diagnosed and treated, and exited the prostate cancer clinical pathway.

- Individuals did not undergo any further investigation or suffer any inconvenience as a result of their imaging study being used.
- The review of the prostatectomy specimens was not to perform further tests on the specimen. The same information as for the routine clinical assessment was acquired. Instead of a descriptive report (as in routine clinical care) the histopathologists' assessment was reported on a reporting template.
- Any data stored or published was and will be fully anonymised and only accessible by the investigators.
- Data stored will remain on a secure computer without any patient identifiers.

Human Tissue Act

Although, the additional histopathology assessment could not be completely anonymised during the assessment of prostatectomy specimens by the histopathologists it was important to ensure the study complied with the Human Tissue Act 2004 [199]. Advice was sought from Dr Ray Lonsdale, histopathologist, to decide whether the study complied with the act, particularly with the below statement

“the researcher is not in possession, and not likely to come into possession of information that identifies the person from whom it has come; and where the material is used for a specific research project approved by a recognised research ethics committee.”

As the histopathologists were part of the direct care team for these patients and their reports of their assessment of the prostatectomy specimens was documented on reporting proforma without any patient identifiable data, the study complied with the Human Tissue Act.

Timeline and correspondence

The study was submitted via the Integrated Research Application System (IRAS) on 21/12/2015.

The study was submitted under the name Focused Diffusion-Weighted-Imaging in Prostate Cancer under the acronym FODIP.

The IRAS number was 182283. The study was also accepted onto the NIHR clinical portfolio and was given a CPMS ID: 30968.

18th January 2016 - Favourable opinion of proportionate review at Research and Ethics Committee subject to conditions (Appendix 3, page 271)

The committee asked the research team to clarify why it was not possible to blind histopathology to patient identifiers. It was explained that it was possible to blind the histopathologists by using a further researcher to obscure the patient identifiers on each pathology specimen slide, however, this posed a greater risk to the patient namely from the potential to mix slides from different patients. The risk of the pathologists seeing the patient name and hospital number was deemed low, particularly as the histopathologists were part of the direct clinical team.

22nd January 2016 – Favourable opinion of proportionate review at Research and Ethics Committee (Appendix 3, page 275)

13th May 2016 – Notification of non-substantial amendment sent to Health Research Authority

This was to facilitate the transfer of the completely anonymised study images via Image Exchange Portal to Addenbrooke's Hospital secure PACS system to facilitate the reading of the MRI by Rad1. The approval was granted on (Appendix 3, page 277).

Risk to patients

The study team identified level of risks for the study as follows:

Risk to	Severity	Frequency
Patients	Mild discomfort (MRI)	Rare
	Severe adverse event (MRI)	Rare
Study	Low	
Trust	Low	

Chapter 5 Results

Patient descriptive statistics

40 patients were included in the study with a mean age of 64 years, a median of 65 years and a range of 42 to 72 years. The distribution of ages of patients is demonstrated in Figure 5-1 and Shapiro-Wilk test of normality indicated this was not a normal distribution ($p=0.003$). All continuous data within this chapter was tested for normality using the Shapiro-Wilk test and by assessing the histogram. If a normal distribution was observed on the histogram and the p-value of the Shapiro-Wilk was greater than 0.05 then the mean and standard deviation (SD) are described. If the data was not normally distributed (non-bell shaped curve and $p < 0.05$) the median and interquartile ranges (IQR) are described.

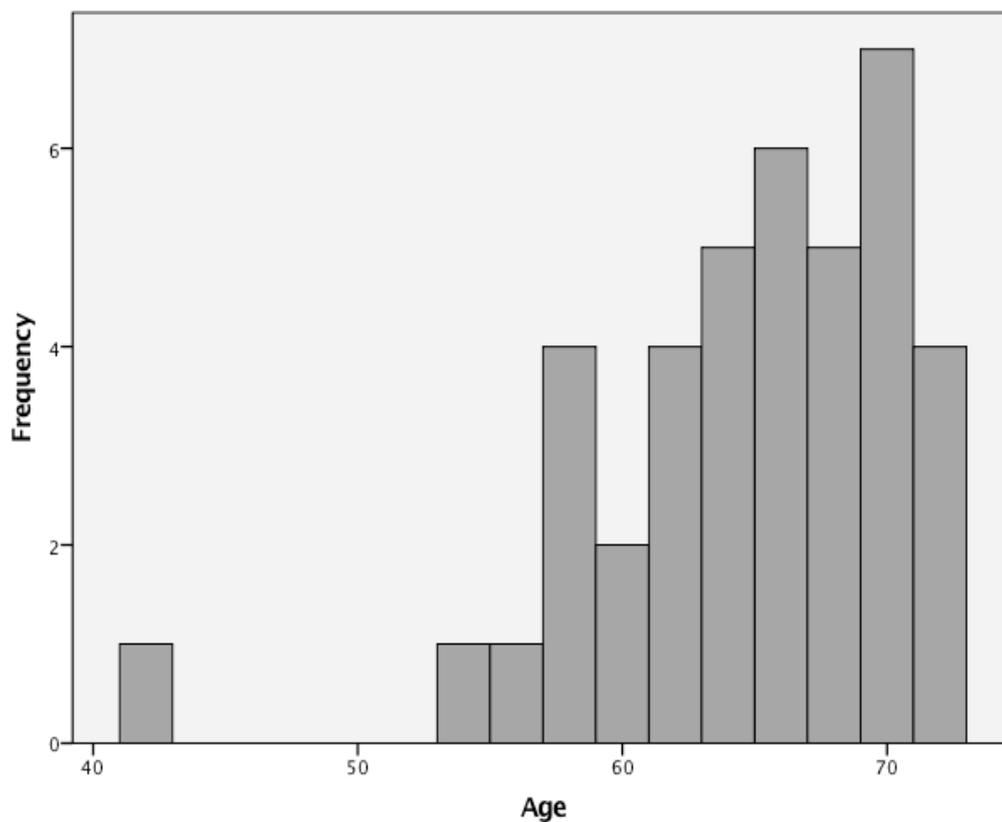


Figure 5-1. Distribution of ages of included patients.

The median duration of biopsy to MRI was 31 days (IQR, 27–39) and the range was -90 days to 406 days (Figure 5-2). Five subjects had their MRI before biopsy, which accounts for the negative numbers, and the remainder had their biopsy after MRI. The subject who waited 406 days for MRI after biopsy had low volume prostate cancer on biopsy and had had a MRI prior to the biopsy, but due to rise in PSA they had a repeat MRI and went on to have a radical prostatectomy.

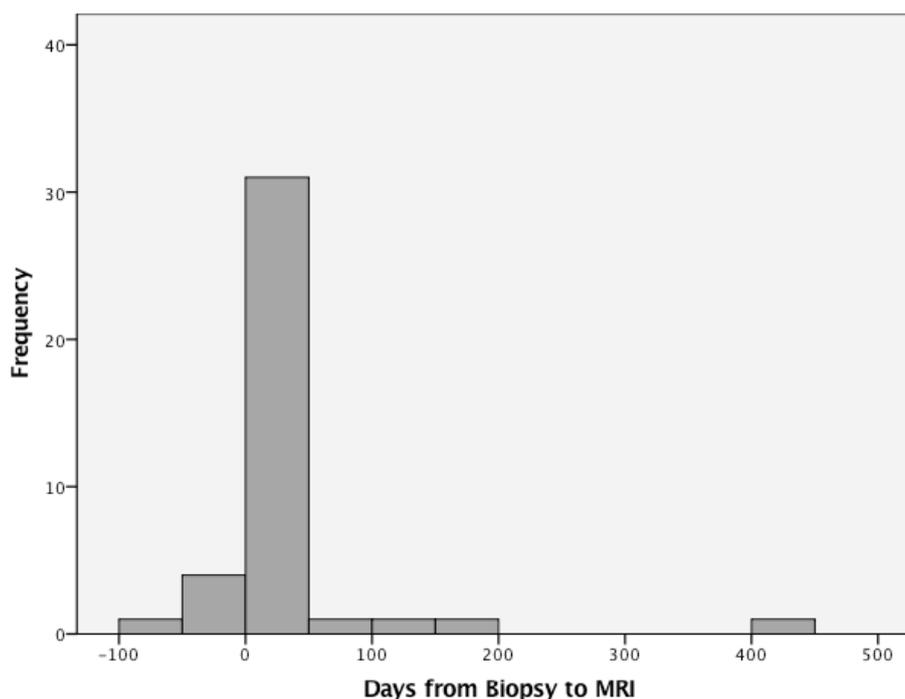


Figure 5-2. Histogram of time from biopsy to MRI.

The median PSA level before biopsy was 8.6 ng/ml (IQR, 6.6–13.4) and the range 4.6 to 34.2 ng/ml. Thirty patients had a TRUS biopsy, nine patients had a Trans-Perineal biopsy and one patient had tumour identified in the pathological specimen obtained at Trans-Urethral Resection of the Prostate. Three subjects had a Gleason Score of 3+3 on biopsy, 25 with 3+4, one with 3+5, nine with 4+3, and two with 4+5 disease. Figure 5-3 demonstrates the distribution of Gleason Grade groups from the biopsy specimens. The median number of days from MRI to radical prostatectomy was 57 days (IQR, 38–75) with a range of 24 to 164 days.

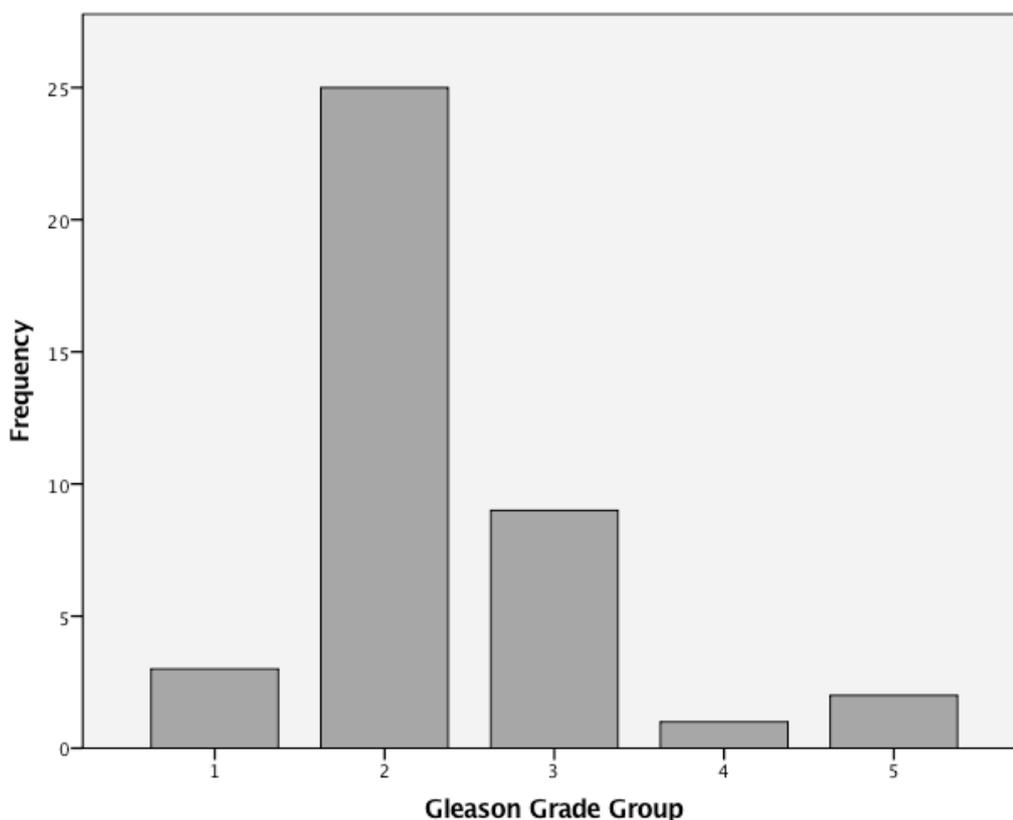


Figure 5-3. Biopsy Gleason Grade Group distribution.

Histopathology descriptive statistics

Of the 40 histopathology specimens 83 lesions were identified with a maximum axial diameter greater than or equal to 5mm. Four separate prostate tumours was the largest number identified in a single prostatectomy specimens. Of the 480 prostate sectors (40 patients with 12 sectors per patient), tumour was present in 238 (49.6%) sectors. In the peripheral zone 166 (69.7%) of 240 sectors contained tumour and in the transitional zone 72 sectors (30.3%) contained tumour.

Post-biopsy haemorrhage, as identified on the T1-weighted images, was present in the location of 21 tumours. The majority of the tumours identified in the radical prostatectomy specimens were Gleason score 3+4=7 and were a Grade Group of 2 (Figure 5-4 and Figure 5-5), but unlike on biopsy specimens there were no Gleason Score 3+3=6 (Grade Group 1) tumours. 79 of the tumours resided predominately in the peripheral zone and 4 in the transitional zone. 17 tumours extended beyond the capsule and three tumours invaded the seminal vesicles. Figure 5-6

demonstrates the increasing likelihood of tumour extension beyond the capsule with increasing Gleason Grade.

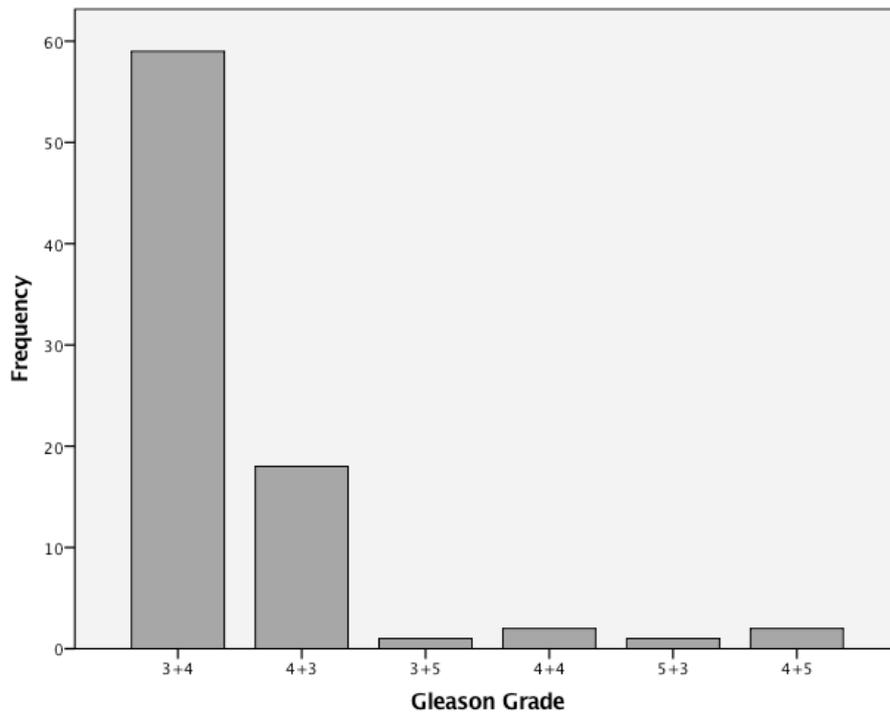


Figure 5-4. Frequencies of tumour Gleason Grade in the included tumours.

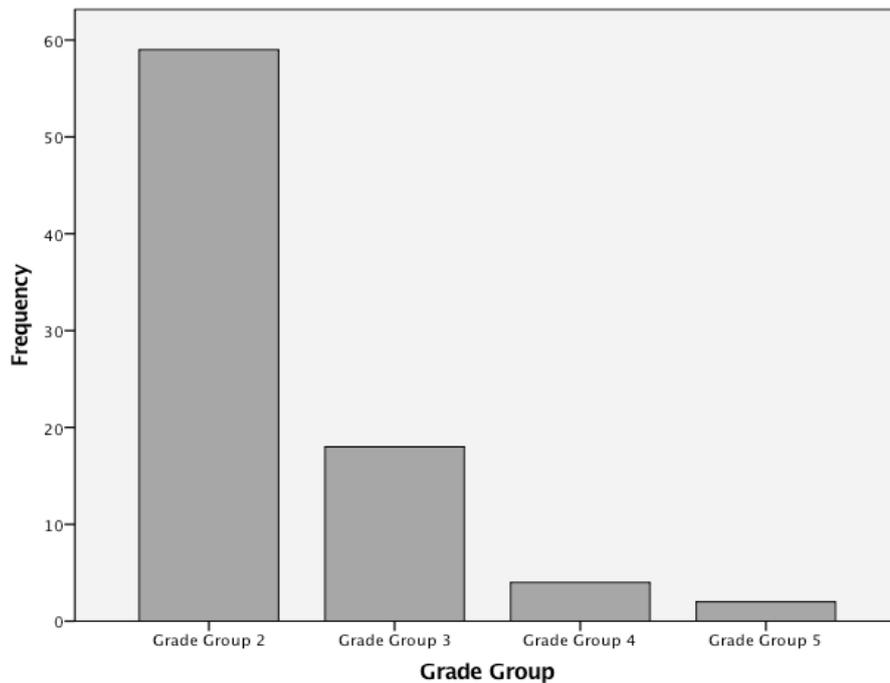


Figure 5-5. Frequencies of tumour Gleason Grade Group in the included tumours.

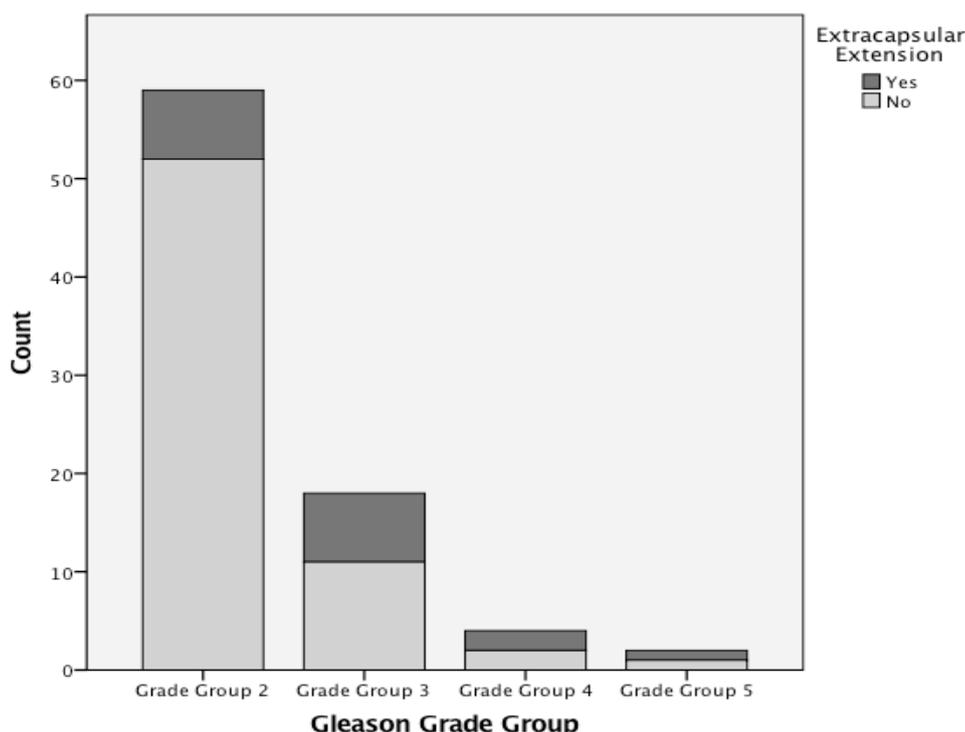


Figure 5-6. Bar chart demonstrating the number of tumours which extended beyond the prostate capsule within different Gleason Grade Groups.

The intraclass correlation coefficient (ICC) measurements for the maximum axial diameter of tumour, and tumour width and height between the two histopathologists were 0.99 (95%CI, 0.99–1.00), 0.99 (95%CI, 0.98–0.99) and 0.90 (95%CI, 0.84–0.94) respectively ($p < 0.001$), indicating “excellent” inter-rater reliability [200,201]. Due to the excellent agreement in tumour size measurements between histopathologists only one (Histopathologist 1) measurement is described.

The median maximum axial diameter was 15 mm (IQR, 11–23 mm) with a range of 5 to 42mm.

The median tumour width and height was 8mm (IQR, 5–11 mm) and 15 mm (IQR, 10–25 mm). The median tumour volume with no correction factor was 1.05cm³ (IQR, 0.32–3.34 cm³).

The median maximum axial diameter for tumours of Gleason Grade Group 2 to 5 are 15mm (IQR, 10–23 mm), 18mm (IQR, 8–23 mm), 16mm (IQR, 12.5–26 mm), and 22 (IQR, 14–30 mm), respectively (Figure 5-7). There was no correlation between maximum axial diameter and Gleason Grade Group as indicated by the Spearman Rank correlation coefficient of 0.058 ($p=0.604$). The

median volume of tumours of Gleason Grade Group tumours 2 to 5 are 0.82cm^3 (IQR, $0.28\text{--}3.34\text{cm}^3$), 1.67cm^3 (IQR, $0.26\text{--}3.47\text{cm}^3$), 1.68cm^3 (IQR, $0.35\text{--}6.16\text{cm}^3$), 2.14 (IQR, $1.17\text{--}3.11\text{cm}^3$), respectively (Figure 5-8). There was no correlation between maximum axial diameter and Gleason Grade Group as indicated by the Spearman Rank correlation coefficient of 0.046 ($p=0.683$).

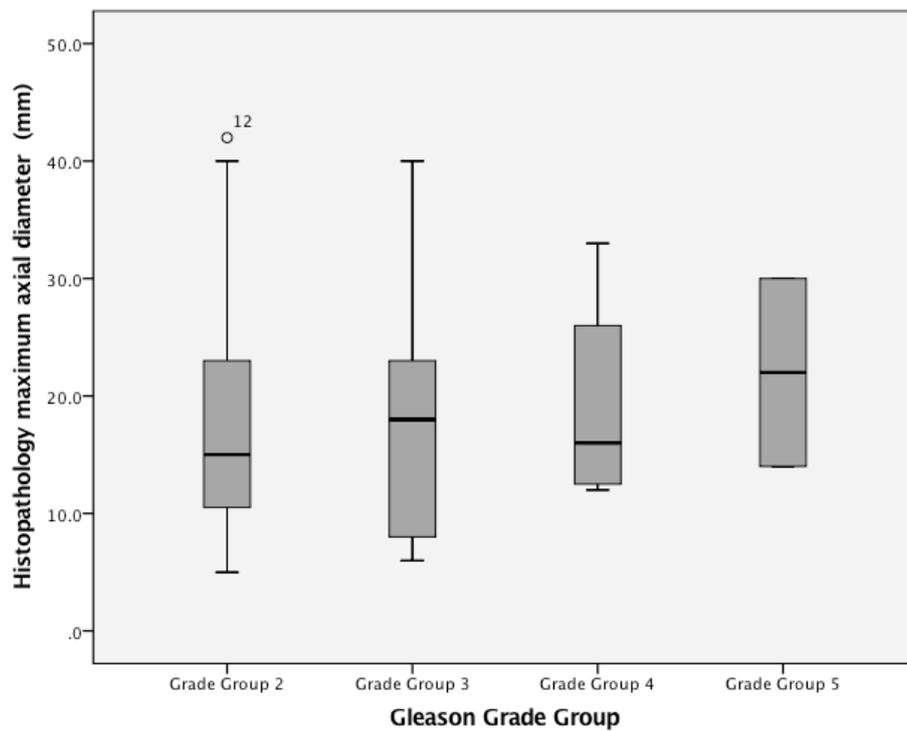


Figure 5-7. Boxplot of maximum axial diameter of different tumour Gleason Grade Groups.

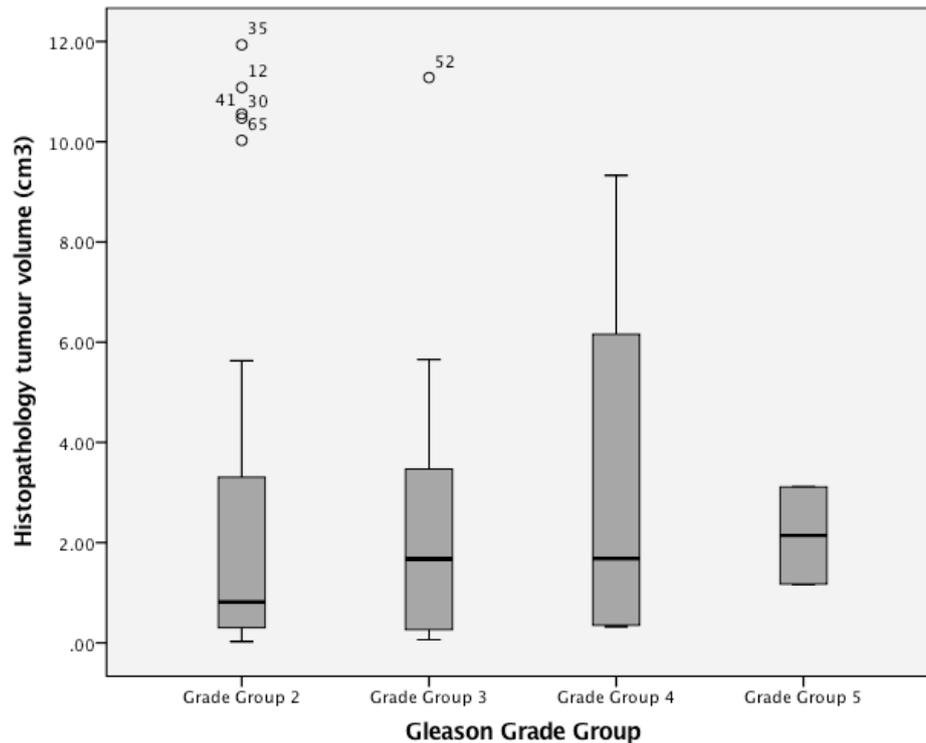


Figure 5-8. Boxplot of tumour volume of different tumour Gleason Grade Groups.

The mean axial diameter of index tumours was 23.0 ± 9.69 mm and the median axial diameter of non-index tumours was 12 mm (IQR, 8–17 mm), which was a statistically significant difference ($p < 0.001$, Mann-Whitney U test) (Figure 5-9). The median tumour volume of index tumours was 3.10 cm^3 (IQR, $0.61\text{--}5.14 \text{ cm}^3$) and non-index tumours was 0.44 cm^3 (IQR, $0.11\text{--}1.17 \text{ cm}^3$) which was also significantly different ($p < 0.001$, Mann-Whitney U test) (Figure 5-10).

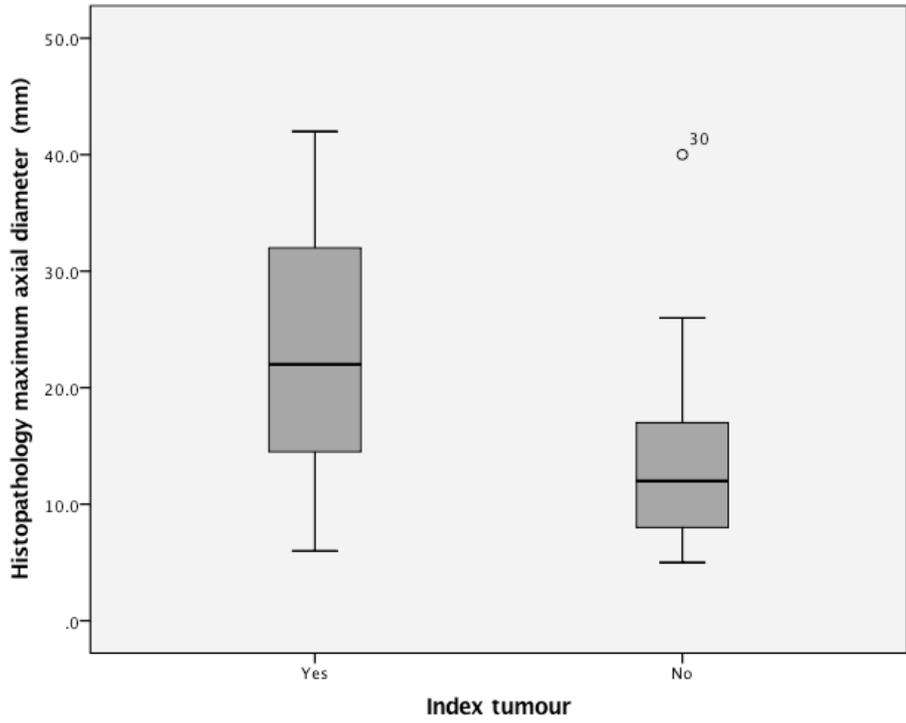


Figure 5-9. Comparison of tumour diameter of index and non-index lesions.

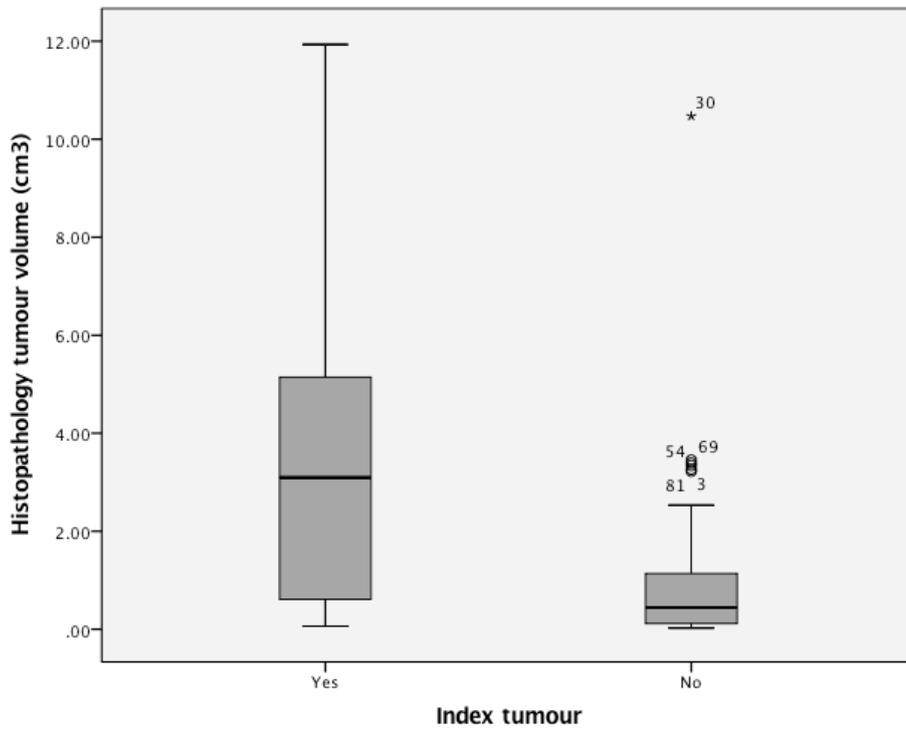


Figure 5-10. Comparison of tumour volume of index and non-index lesions.

The median maximum axial diameter of tumours confined by the capsule was 13mm (IQR, 10–22 mm) and mean diameter of tumours which extended beyond the capsule was 25.1 ± 8.69 mm, a significant difference ($p=0.001$, Mann Whitney U test) (Figure 5-11). There was a significance difference between the tumour volume of tumours which extended beyond the capsule (median, 3.10cm^3 ; IQR, $0.61\text{--}5.14\text{cm}^3$) and those which did not (median, 0.44cm^3 ; IQR, $0.11\text{--}1.17\text{cm}^3$) ($p < 0.001$, Mann Whitney U test) (Figure 5-12).

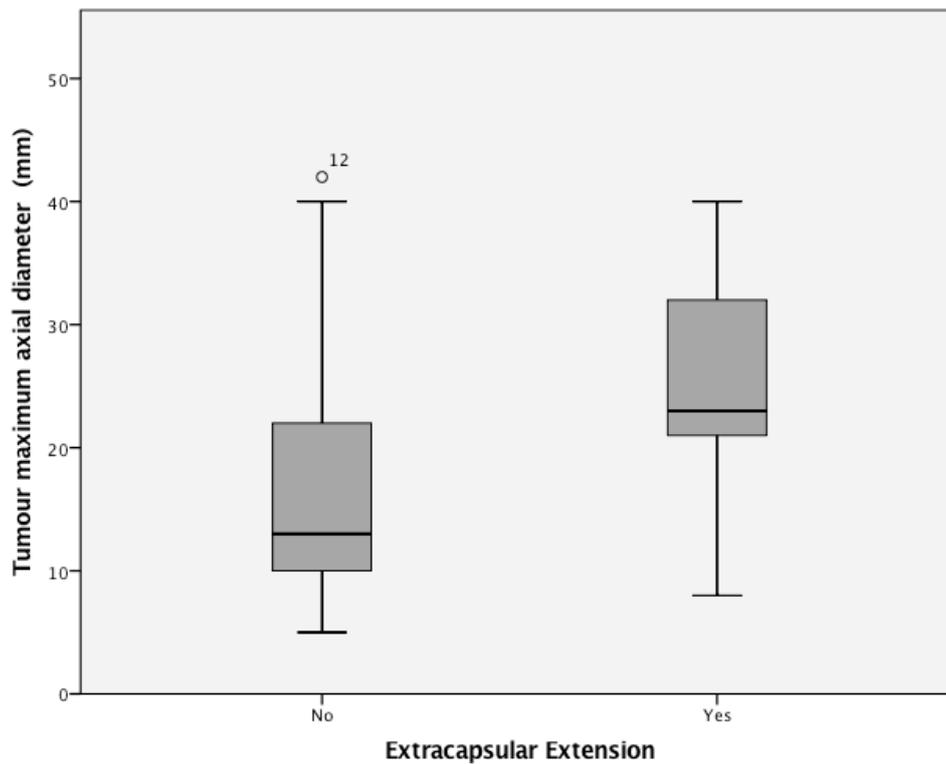


Figure 5-11. Comparison of tumour diameter of tumours confined to and those which extended beyond the capsule.

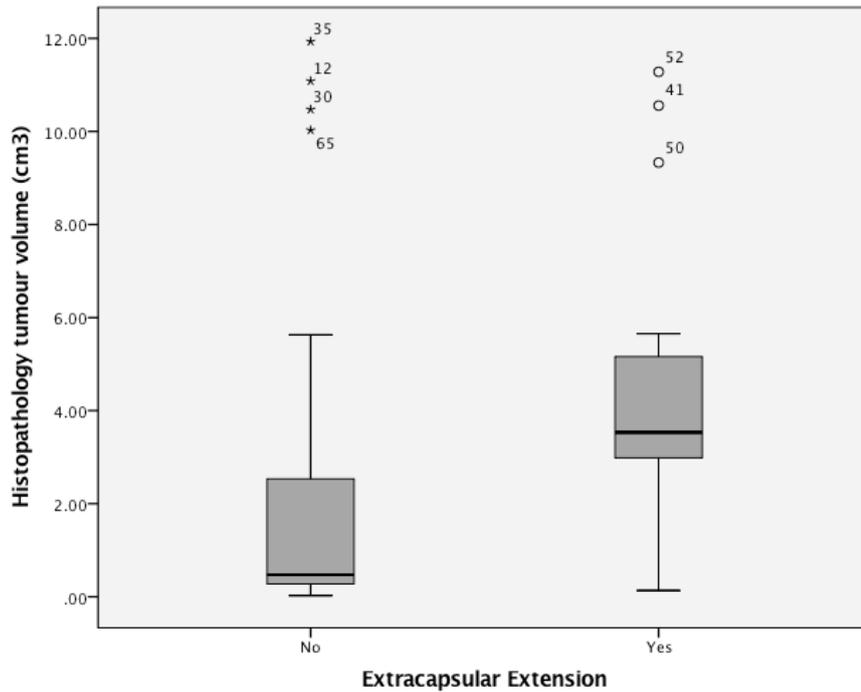


Figure 5-12. Boxplot of tumour volume of tumours which extended beyond the capsule and those which did not.

Three tumours invaded the seminal vesicles. The median axial diameter of tumours that did not invade the seminal vesicles was 15mm (IQR, 11–23 mm), and the mean diameter of tumours which did invade the seminal vesicles was 20 ± 12.5 mm ($p = 0.791$, Mann-Whitney U) (Figure 5-13).

There was no significant difference ($p=0.635$) between the tumour volumes of tumours which did and did not invade the seminal vesicles and shown in Figure 5-14.

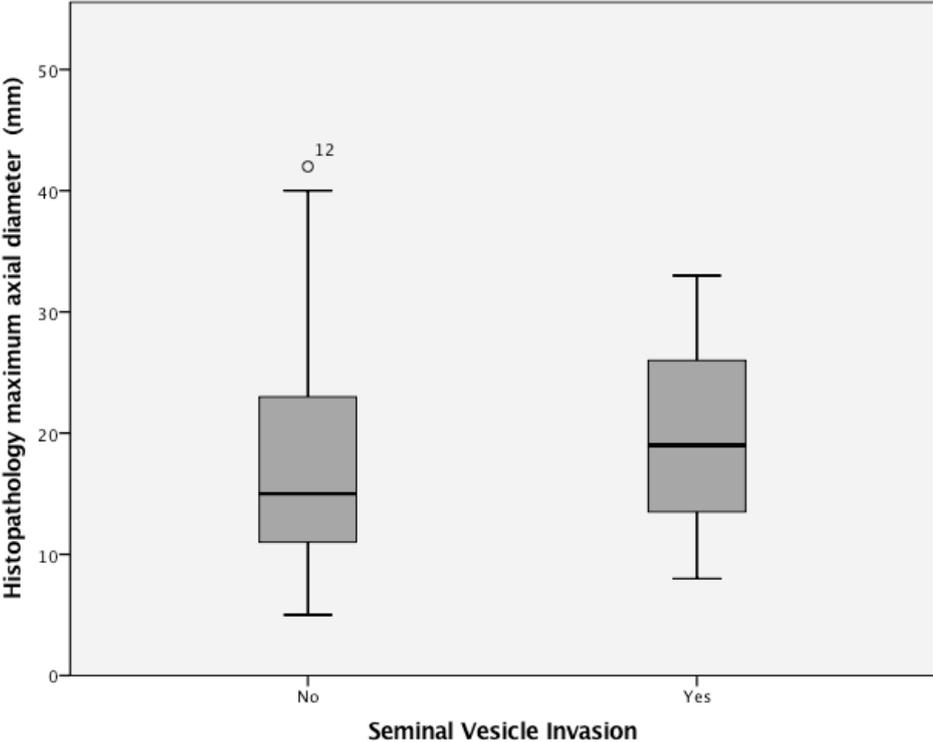


Figure 5-13. Boxplot displaying the difference in maximum axial diameters of tumours which did and did not invade the seminal vesicles.

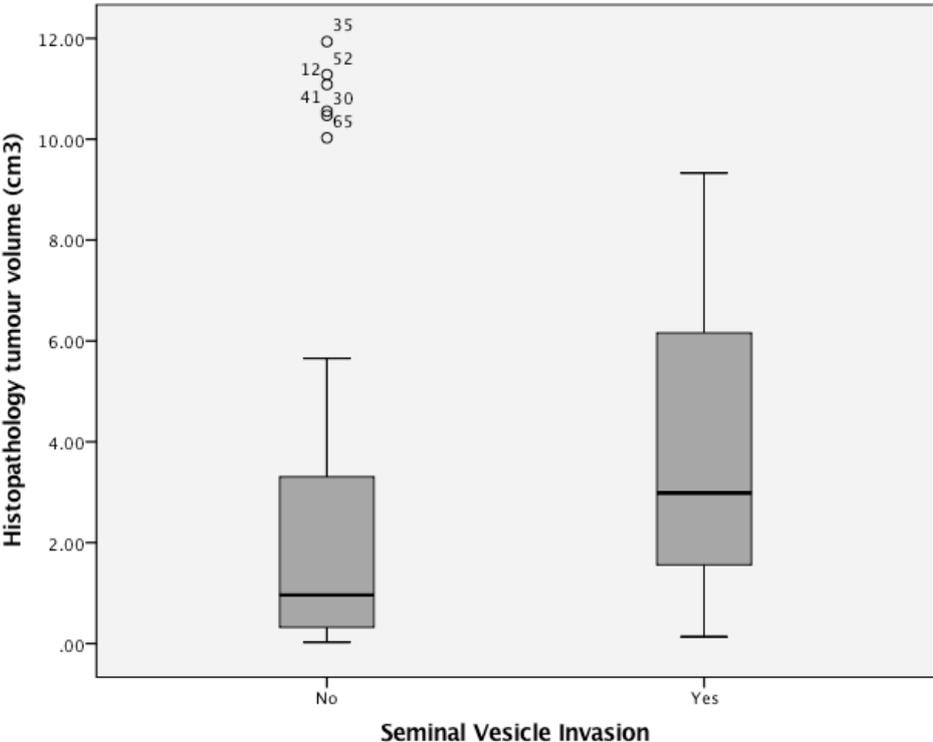


Figure 5-14. Boxplot of tumour volume of lesions which did and did not invade the seminal vesicles.

Diagnostic performance of DWI sequences: sector-based analysis

The apex of the prostate was not covered on the sDWI sequence in two cases, as a result the 4 apical sectors of these cases were omitted from the analysis and thus 476 sectors were analysed in the sDWI (protocol C) and sDWI + T2WI (protocol D) reads. The cDWI analysis included 12 sectors in all 40 cases, with 480 sectors in cDWI (protocol A) and cDWI + T2WI (protocol B) reads.

Both readers regardless of imaging protocol had more Likert scores of 1 and 5 compared to scores of 2 to 4, and reader 2 had few Likert scores of 2 (Table 5-1 and 5.2). Both readers, generally, correctly identified when tumour was absent from a sector and when tumour was present with all imaging protocols (Figure 5-15 A-D and Figure 5-16 A-D).

Table 5-1. Reader 1 frequency of Likert scores in different protocols.

Reader 1	Protocol			
Likert score	cDWI (A)	cDWI + T2WI (B)	sDWI (C)	sDWI + T2WI (D)
Tumour definitely absent (1)	206	214	262	244
Tumour probably absent (2)	90	87	82	95
Indeterminate (3)	26	20	22	18
Tumour probably present (4)	23	27	14	14
Tumour definitely present (5)	135	132	88	101

Table 5-2. Reader 2 frequency of Likert scores in different protocols.

Reader 2	Protocol			
Likert score	cDWI (A)	cDWI + T2WI (B)	sDWI (C)	sDWI + T2WI (D)
Tumour definitely absent (1)	319	305	333	310
Tumour probably absent (2)	7	7	8	13
Indeterminate (3)	18	11	16	4
Tumour probably present (4)	51	20	20	32
Tumour definitely present (5)	85	137	95	113

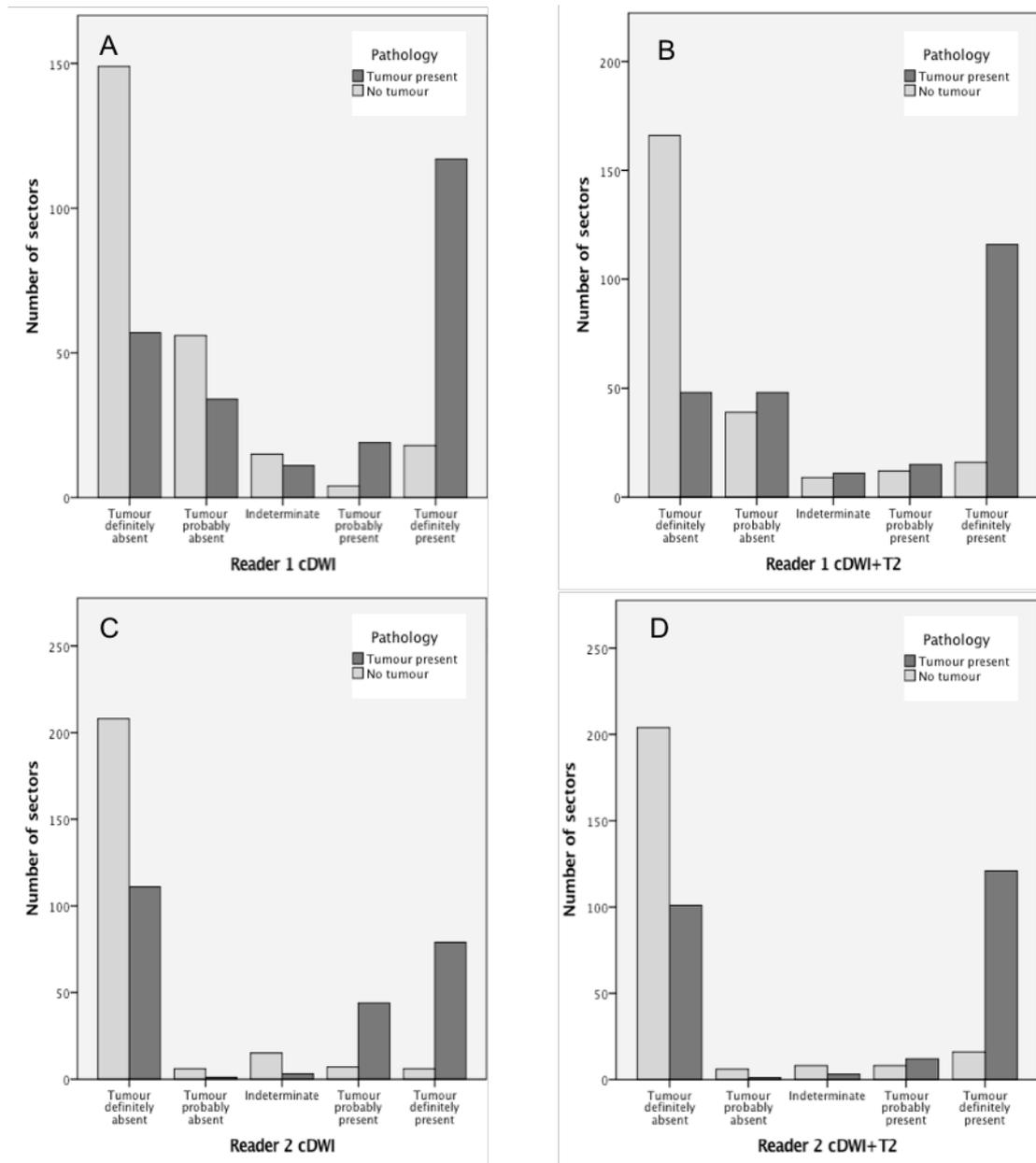


Figure 5-15. Bar chart demonstrating the ratio of sectors with tumour present or absent to probability of the tumour presence with the conventional DWI sequence. A – Reader 1 Protocol A, B – Reader 1 Protocol B, C – Reader 2 Protocol A, D - Reader 2 Protocol B.

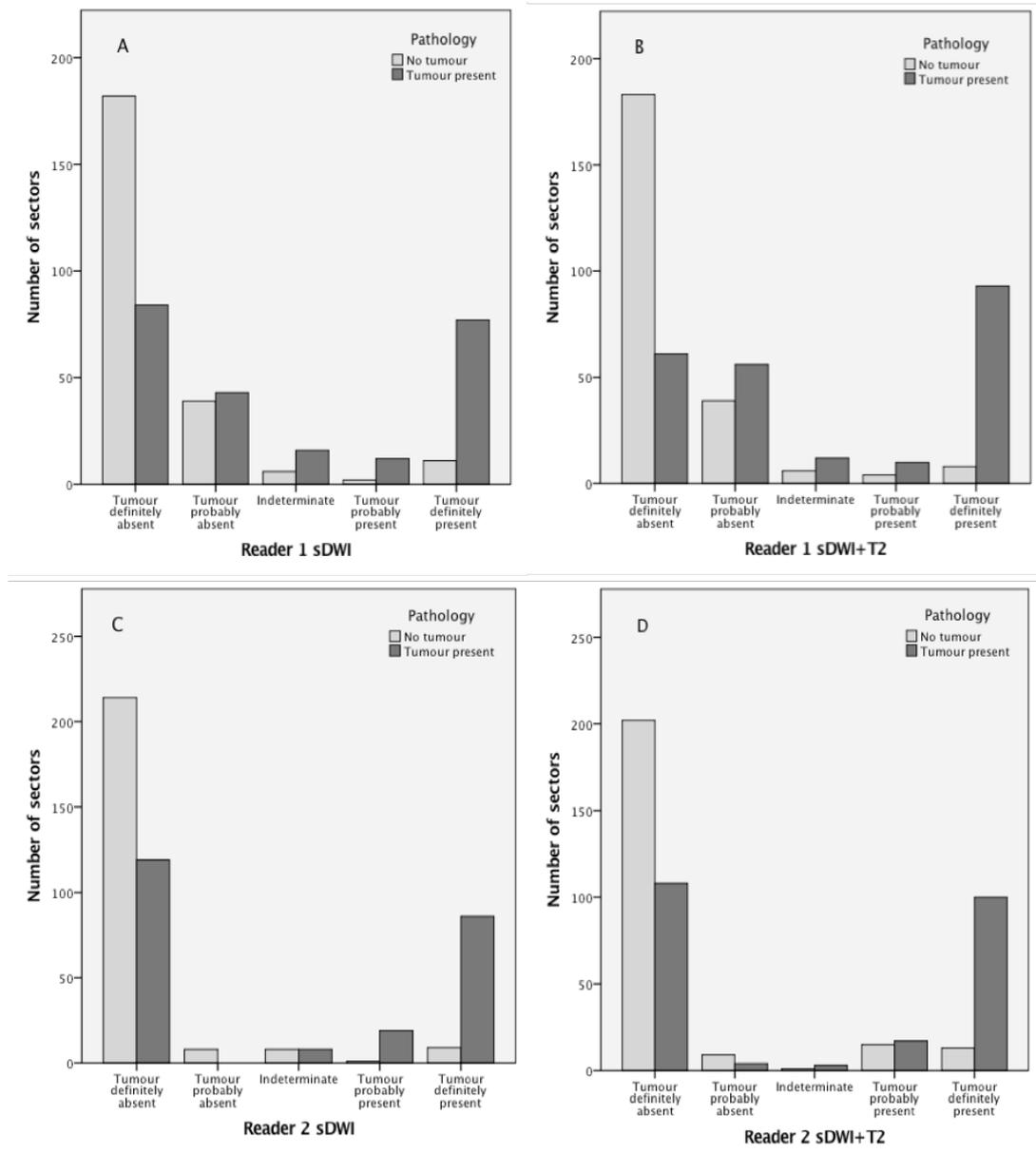


Figure 5-16. Bar chart demonstrating the ratio of sectors with tumour present or absent to probability of the tumour presence with the small FOV DWI sequence. A - Reader 1 Protocol A, B - Reader 1 Protocol B, C - Reader 2 Protocol A, D - Reader 2 Protocol B.

The 2 x 2 contingency table was completed for both readers and all protocols using a Likert score 3 and greater as positive for tumour and also 4 and greater. A threshold effect was demonstrated throughout with less true and false positive results and more true and false negative results with the higher Likert cut-off. The diagnostic results of each protocol for both readers are presented in Table 5-3 and Table 5-4.

Table 5-3. Diagnostic performance of reader 1 for each imaging protocol on a per sector basis.

Reader 1	Likert cut-off	True Positives	False Positives	True Negatives	False Negatives
cDWI (Protocol A)	3	147	37	205	91
	4	136	22	220	102
cDWI + T2WI (Protocol B)	3	142	37	205	96
	4	131	28	214	107
sDWI (Protocol C)	3	105	19	221	127
	4	89	13	227	143
sDWI + T2WI (Protocol D)	3	115	18	222	117
	4	103	12	228	129

Table 5-4. Diagnostic performance of reader 2 for each imaging protocol on a per sector basis.

Reader 2	Likert cut-off	True Positives	False Positives	True Negatives	False Negatives
cDWI (Protocol A)	3	126	28	214	112
	4	123	13	229	115
cDWI + T2WI (Protocol B)	3	136	32	210	102
	4	133	24	218	105
sDWI (Protocol C)	3	113	18	222	119
	4	105	10	230	127
sDWI + T2WI (Protocol D)	3	120	29	211	112
	4	117	28	212	115

The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) with 95% CI for Likert cut-off's of 3 and 4 are presented in Table 5-5 and Table 5-6. The statistical comparison of the comparable sDWI and cDWI imaging protocols for each of the diagnostic parameters is also shown.

For reader 1 there was a significant improvement in sensitivity of the cDWI sequence over sDWI regardless of the presence of the T2WI sequence (Likert 4 cut-off, $p = <0.001$), and conversely the sDWI specificity was significantly better than cDWI (Likert cut-off 3, $p = 0.003$). Reader 2 found no significant difference in sensitivity or specificity between cDWI and sDWI regardless of cut-off or protocol. The sDWI sequence sensitivity significantly improved when the T2WI sequence was added for reader 1 (Likert 4 cut-off, $p = 0.007$) and the same was observed with reader 2 for the cDWI sequence (Likert cut-off 4, $p = <0.001$). The specificity was significantly lower for reader 2 when T2WI was added to either the cDWI (Likert cut-off 4, $p = 0.001$) and sDWI sequence (Likert cut-off 4, $p = <0.001$). There was little difference between the PPV and NPV for reader 1 and reader 2 when sDWI was compared to cDWI. McNemar test was performed to compare the sensitivity and specificity results, and a test of proportions was performed for PPV and NPV

Table 5-5. Diagnostic performance of reader 1 and comparison of cDWI and sDWI sequences.

Reader 1	Likert cut-off	Protocol A cDWI	Protocol C sDWI	Protocol B cDWI + T2WI	Protocol D sDWI + T2WI	A vs C (p-value)	B vs D (p-value)	A vs B (p-value)	C vs D (p-value)
Sensitivity	3	0.62 (0.55-0.68)	0.45 (0.39-0.52)	0.60 (0.53-0.66)	0.50 (0.43-0.56)	<0.001*	0.007	0.499	0.041
	4	0.57 (0.51-0.63)	0.38 (0.32-0.45)	0.55 (0.48-0.61)	0.44 (0.38-0.51)	<0.001*	0.005	0.441	0.007
Specificity	3	0.85 (0.80-0.89)	0.92 (0.88-0.95)	0.85 (0.80-0.89)	0.93 (0.88-0.95)	0.003*	0.004	1.000	1.000
	4	0.91 (0.87-0.94)	0.95 (0.91-0.97)	0.88 (0.84-0.92)	0.95 (0.91-0.97)	0.096	0.011	0.263	1.000
PPV	3	0.80 (0.73-0.85)	0.85 (0.77-0.91)	0.79 (0.73-0.85)	0.86 (0.79-0.92)	0.285	0.102	0.894	0.683
	4	0.86 (0.80-0.91)	0.87 (0.79-0.93)	0.82 (0.76-0.88)	0.90 (0.82-0.94)	0.786	0.097	0.368	0.595
NPV	3	0.69 (0.64-0.74)	0.64 (0.58-0.69)	0.68 (0.63-0.73)	0.65 (0.60-0.71)	0.124	0.483	0.762	0.588
	4	0.68 (0.63-0.73)	0.61 (0.56-0.66)	0.67 (0.61-0.72)	0.64 (0.59-0.69)	0.056	0.445	0.654	0.484
AUC		0.76 (0.72-0.81)	0.73 (0.69-0.77)	0.79 (0.75-0.83)	0.79 (0.75-0.83)	0.164	0.818	0.138	<0.001

* Denotes the comparison remains significantly different following Bonferroni correction for multiple comparisons.

Table 5-6. Diagnostic performance of reader 2 and comparison of cDWI and sDWI sequences.

Reader 2	Likert cut-off	Protocol A cDWI	Protocol C sDWI	Protocol B cDWI + T2WI	Protocol D sDWI + T2WI	A vs C (p-value)	B vs D (p-value)	A vs B (p-value)	C vs D (p-value)
Sensitivity	3	0.53 (0.46-0.60)	0.49 (0.42-0.55)	0.57 (0.51-0.64)	0.52 (0.45-0.58)	0.403	0.170	0.002*	0.189
	4	0.52 (0.45-0.58)	0.45 (0.39-0.52)	0.56 (0.49-0.62)	0.50 (0.44-0.57)	0.143	0.161	0.002*	0.012
Specificity	3	0.88 (0.84-0.92)	0.93 (0.88-0.95)	0.87 (0.82-0.91)	0.88 (0.83-0.92)	0.112	0.719	0.388	0.019
	4	0.95 (0.91-0.97)	0.96 (0.93-0.98)	0.90 (0.86-0.94)	0.88 (0.84-0.92)	0.581	0.571	0.001*	<0.001*
PPV	3	0.82 (0.75-0.88)	0.86 (0.79-0.92)	0.81 (0.74-0.87)	0.81 (0.73-0.87)	0.310	0.977	0.842	0.201
	4	0.90 (0.84-0.95)	0.91 (0.85-0.96)	0.85 (0.78-0.90)	0.81 (0.73-0.87)	0.813	0.355	0.141	0.016
NPV	3	0.66 (0.60-0.71)	0.65 (0.60-0.70)	0.67 (0.62-0.72)	0.65 (0.60-0.71)	0.883	0.883	0.656	0.952
	4	0.67 (0.61-0.72)	0.64 (0.59- 0.69)	0.67 (0.62-0.73)	0.65 (0.59-0.70)	0.551	0.474	0.800	0.912
AUC		0.72 (0.68-0.76)	0.70 (0.66-0.74)	0.73 (0.69-0.77)	0.71 (0.67-0.75)	0.547	0.318	0.265	0.675

* Denotes the comparison remains significantly different following Bonferroni correction for multiple comparisons.

The ROC for reader 1 are displayed in Figure 5-17, and for reader 2 in Figure 5-18. For protocols A and C the cDWI AUC was higher than sDWI, but the difference was not significant (DeLong’s test) . The only significant difference ($p = <0.001$) was the significant increase in AUC for reader 1’s sDWI analysis when T2WI was added to the assessment in (Table 5-5 and Table 5-6).

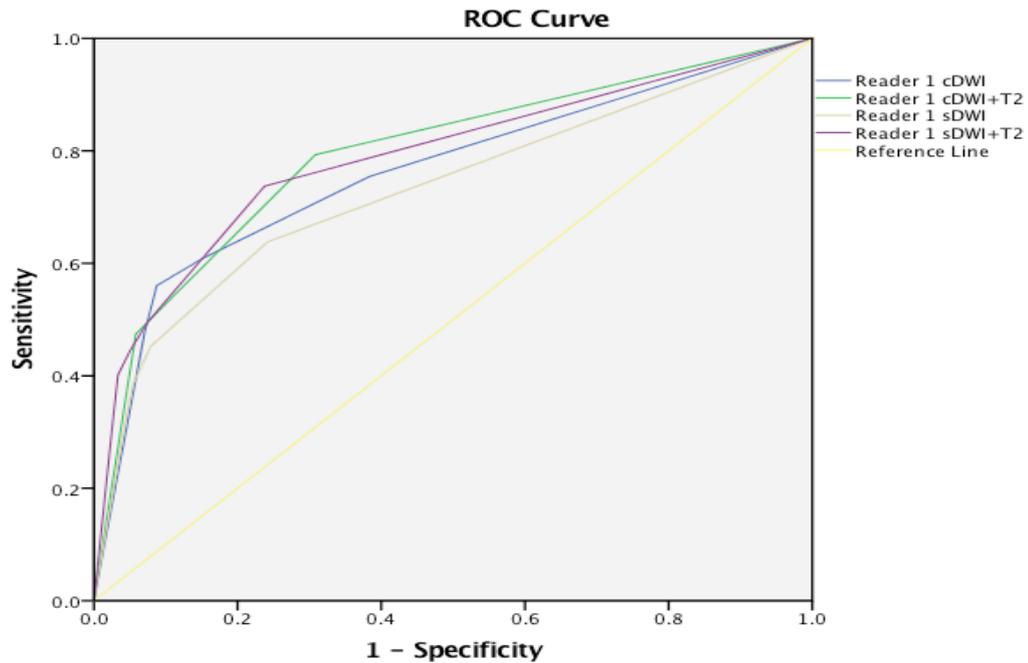


Figure 5-17. Receiver operating characteristic curve of sectoral diagnostic performance of reader 1.

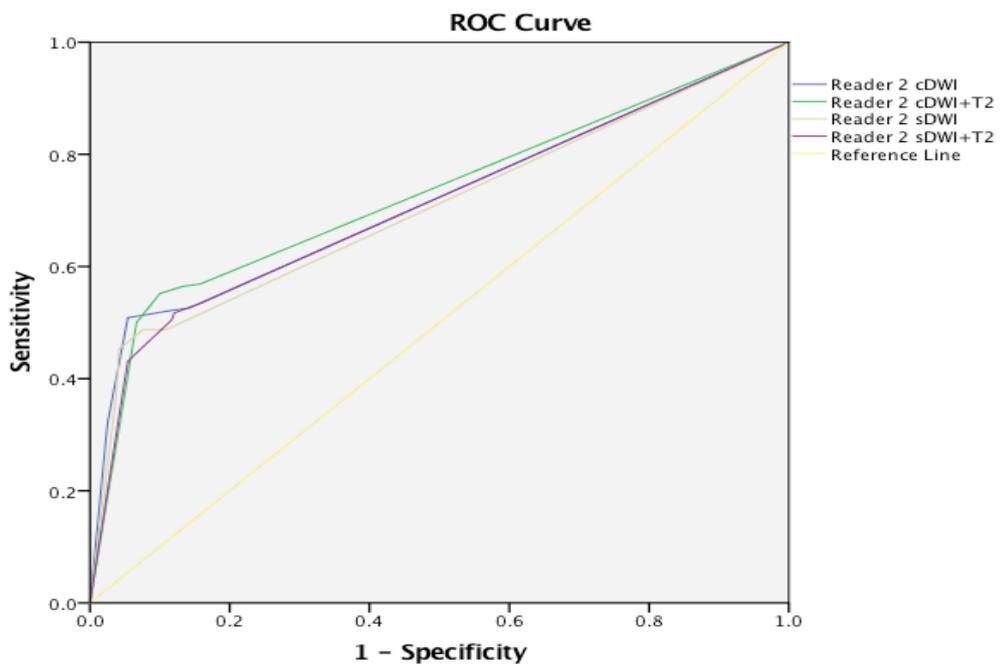


Figure 5-18. Receiver operating characteristic curve of sectoral diagnostic performance of reader 2.

Sectoral reliability

Interobserver agreement of reader 1 and 2 of the sectoral analysis as denoted by the weighted kappa (κ) with 95% CI is demonstrated in Table 5-7. The weighted kappa was not significantly different between cDWI and sDWI with and without T2WI (two proportion Z-test). The interobserver reliability was not significantly different between the peripheral and transitional zones within each sequence. Agreement dropped after the addition of the T2WI sequence, but the difference was not significant. The interobserver agreement was substantial for cDWI sequence and moderate for the remainder of the sequences [200].

Table 5-7. Interobserver reliability, as demonstrated by weighted kappa tests, between reader 1 and 2 for all sectors, and the peripheral zone and transitional zone sectors.

Sectors	cDWI (A)	sDWI (C)	A vs C (p-value)	cDWI + T2WI (B)	sDWI + T2WI (D)	B vs D (p-value)
All	0.62 (0.55–0.69)	0.59 (0.51–0.67)	0.58	0.54 (0.47–0.62)	0.52 (0.43–0.60)	0.74
Peripheral zone	0.62 (0.51–0.72)	0.59 (0.47–0.70)	0.72	0.56 (0.45–0.67)	0.50 (0.38–0.62)	0.48
Transitional zone	0.62 (0.52–0.71)	0.58 (0.48–0.70)	0.58	0.52 (0.42–0.63)	0.52 (0.41–0.63)	0.99

Diagnostic performance of DWI sequences: lesion-based analysis

83 lesions were identified on histopathological analysis of the 40 patients' radical prostatectomy specimens. Two lesions were excluded from the sDWI MRI reads (Protocols C and D) as the apex and lesions were not included on the images.

In the comparable sequences both reader 1 and 2 identified more index lesions on cDWI than sDWI, with and without T2WI, as demonstrated in Table 5-8. The best tumour detection rate was 87.5%, which was in reader 1's clinical assessment of the cDWI sequence without T2WI.

Table 5-8. Index lesion identification of reader 1 and 2 with all imaging protocol following a stringent and clinical assessment of readers' results.

	Protocol	Stringent		Clinical	
		Seen	%	Seen	%
Reader 1	cDWI (A)	33	82.5	35	87.5
	sDWI (C)	30	75	33	82.5
	cDWI + T2WI (B)	33	82.5	34	85
	sDWI + T2WI (D)	31	77.5	33	82.5
Reader 2	cDWI (A)	31	77.5	34	85
	sDWI (C)	30	75	31	77.5
	cDWI + T2WI (B)	33	82.5	34	85
	sDWI + T2WI (D)	32	80	32	80

The positive predictive value of reader 1's lesion assessment was consistently higher with sDWI than cDWI with and without T2WI, and following both 'stringent' and 'clinical' evaluation, which was due to the reduced number of false positive lesions (Table 5-9). The sensitivity was better with cDWI than sDWI without T2WI, but similar with T2WI.

Table 5-9. Reader 1 lesion accuracy.

Reader 1	Protocol	True Positives	False Positives	False Negatives	Sensitivity	PPV
Stringent	cDWI (A)	47	13	36	0.57 (0.45–0.67)	0.78 (0.83–0.99)
	sDWI (C)	42	4	39	0.52 (0.40–0.63)	0.91 (0.79–0.98)
	cDWI + T2WI (B)	47	8	36	0.57 (0.45–0.67)	0.86 (0.71–0.95)
	sDWI + T2WI (D)	46	4	35	0.57 (0.45–0.68)	0.92 (0.81–0.98)
Clinical	cDWI (A)	53	13	30	0.64 (0.53–0.74)	0.80 (0.69–0.89)
	sDWI (C)	46	4	35	0.57 (0.45–0.68)	0.92 (0.81–0.98)
	cDWI + T2WI (B)	50	8	33	0.60 (0.49–0.71)	0.86 (0.75–0.94)
	sDWI + T2WI (D)	50	4	31	0.62 (0.50–0.72)	0.93 (0.82–0.98)

Similar findings were observed in reader 2's lesion accuracy assessment (Table 5-10). cDWI was more sensitive than sDWI with and without T2WI, but the reverse was seen with positive predictive value and a reduced number of false positives. A maximum sensitivity of 59% is observed with cDWI and a maximum positive predictive value of 98% was demonstrated with sDWI and T2WI combined.

Table 5-10. Reader 2 lesion accuracy.

Reader 2	Protocol	True Positive	False Positives	False Negatives	Sensitivity	PPV
Stringent	cDWI	43	5	40	0.52 (0.41–0.63)	0.90 (0.77–0.97)
	sDWI	40	2	41	0.49 (0.38–0.61)	0.95 (0.84–0.99)
	cDWI + T2WI	46	7	43	0.55 (0.44–0.66)	0.87 (0.75–0.95)
	sDWI + T2WI	41	2	42	0.49 (0.39–0.61)	0.95 (0.84–0.99)
Clinical	cDWI	49	5	34	0.59 (0.48–0.70)	0.91 (0.80–0.97)
	sDWI	43	2	39	0.53 (0.42–0.64)	0.96 (0.85–0.99)
	cDWI + T2WI	49	7	34	0.59 (0.48–0.70)	0.88 (0.76–0.95)
	sDWI + T2WI	43	1	38	0.53 (0.42–0.64)	0.98 (0.88–1.00)

Effects of haemorrhage and tumour grade on lesion-based accuracy

An assessment of the percentage of lesions seen which were affected and not affected by haemorrhage was performed (Figure 5-19 – 5-22). Lesions outwith an area of haemorrhage were detected more frequently by both radiologists, with both DWI sequences and with or without T2WI.

Tumours were more likely to be seen if they were of higher grade regardless of imaging protocol (Figure 5-23 - 5-26).

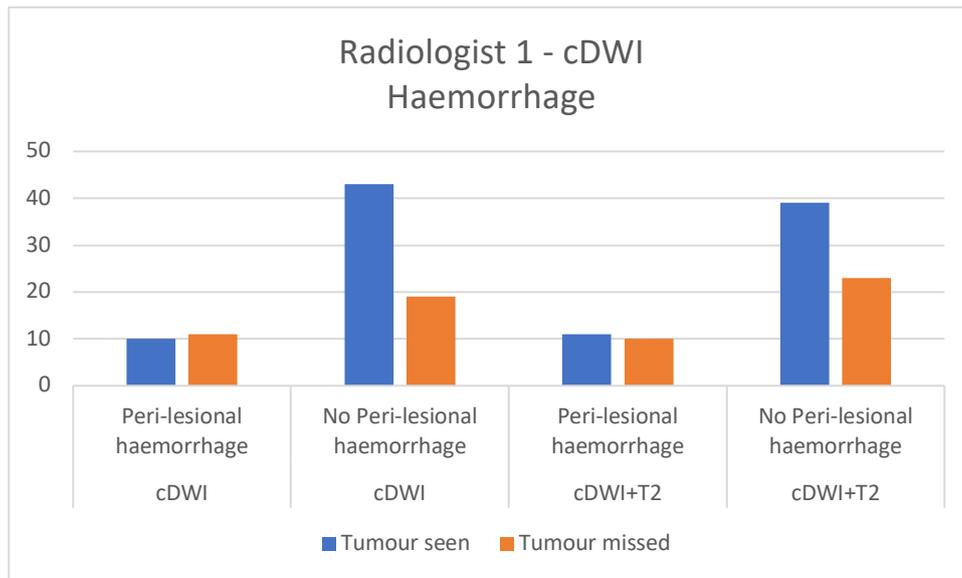


Figure 5-19. Influence of peri-lesional haemorrhage on Radiologist 1's ability to detect and localise tumour with cDWI and T2WI.

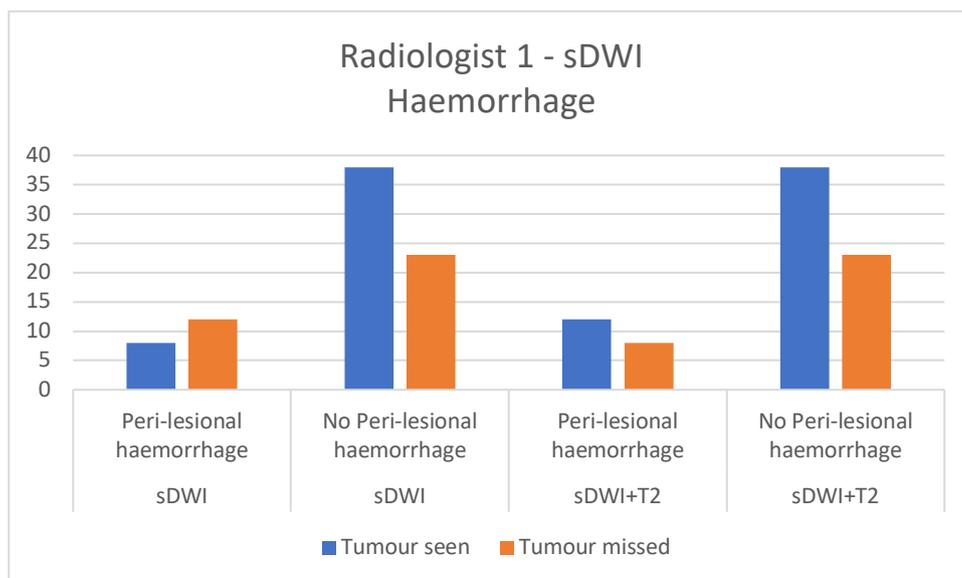


Figure 5-20. Influence of peri-lesional haemorrhage on Radiologist 1's ability to detect and localise tumour with sDWI and T2WI.

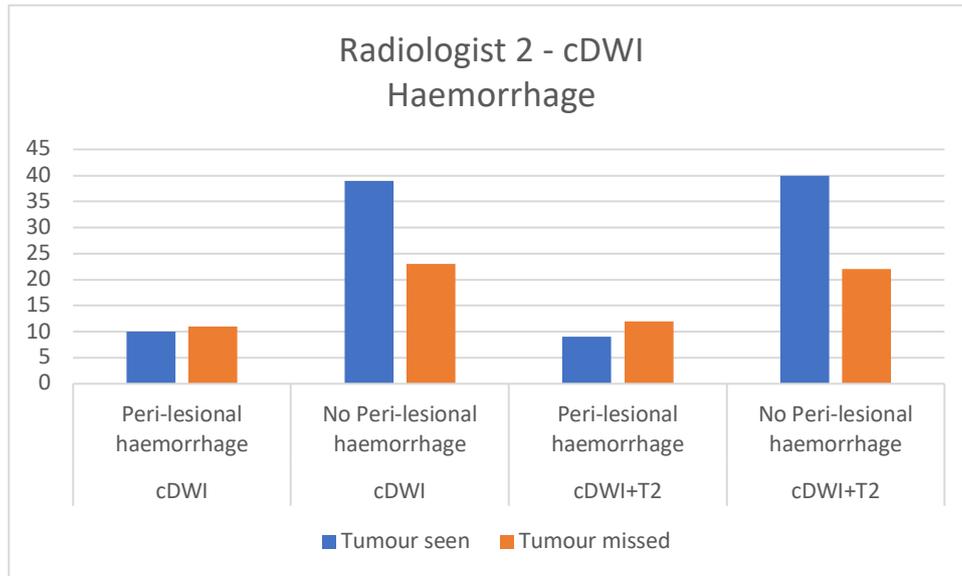


Figure 5-21. Influence of peri-lesional haemorrhage on Radiologist 2's ability to detect and localise tumour with cDWI and T2WI.

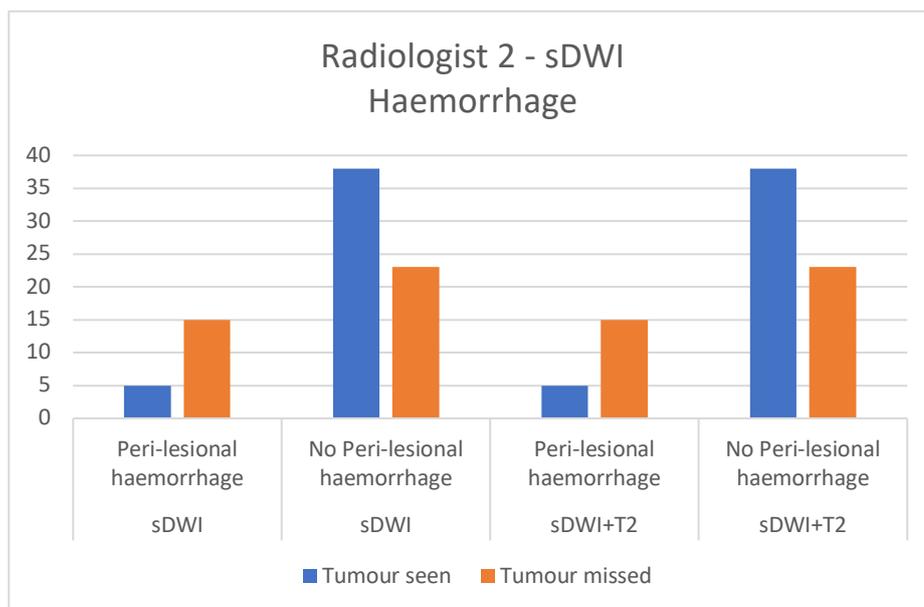


Figure 5-22. Influence of peri-lesional haemorrhage on Radiologist 2's ability to detect and localise tumour with sDWI and T2WI.

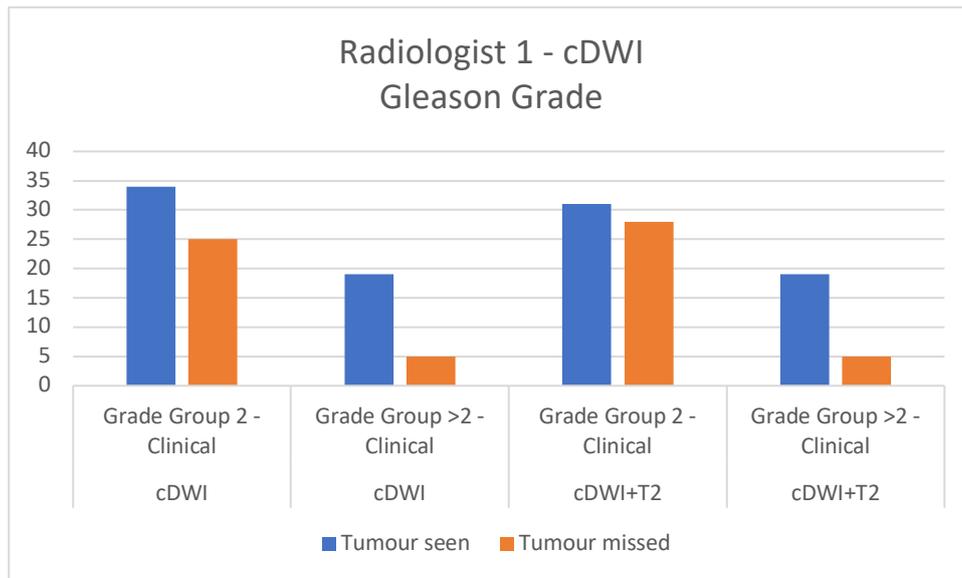


Figure 5-23. Comparison of detection rate of tumours of different Gleason Grade Group with cDWI by Radiologist 1.

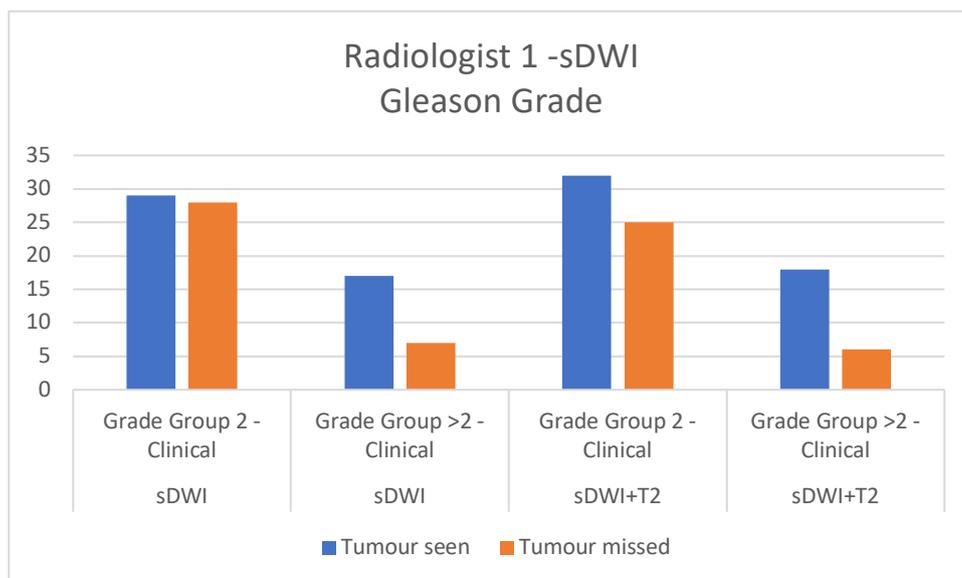


Figure 5-24. Comparison of detection rate of tumours of different Gleason Grade Group with sDWI by Radiologist 1.

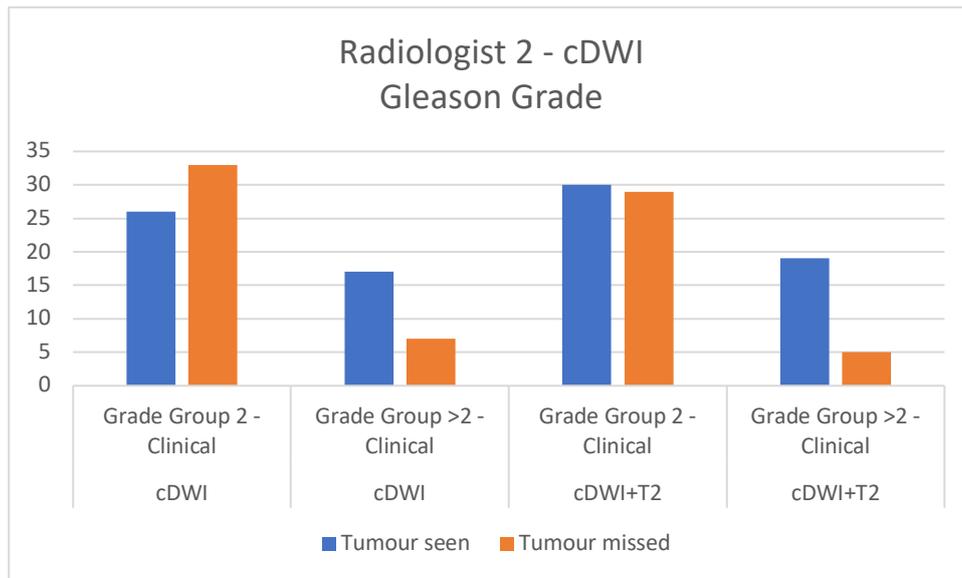


Figure 5-25. Comparison of detection rate of tumours of different Gleason Grade Group with cDWI by Radiologist 1

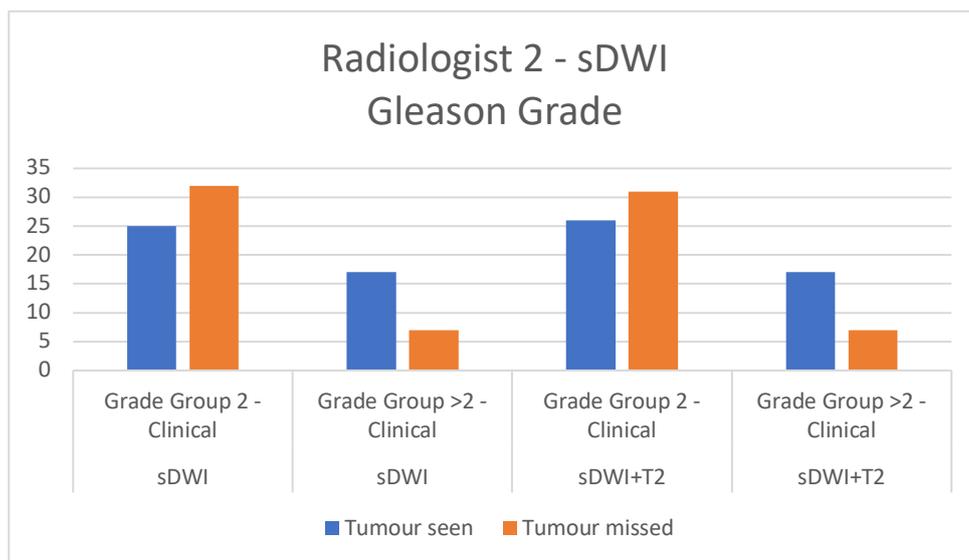


Figure 5-26. Comparison of detection rate of tumours of different Gleason Grade Group with sDWI by Radiologist 2.

Quantitative analysis of lesions

Signal intensity of tumour and non-tumour tissue

The signal intensity (SI) of tumour was lower than non-tumour (normal) tissue on the ADC maps of both DWI sequences and higher than non-tumour tissue on the b1000 and b1500 images (Table 5-11 and Table 5-12). The signal intensity between tumour and non-tumour tissue was significantly different on all sequences except on the b1000 images of the cDWI sequence. The signal intensity measured on ADC maps is also a measure of the ADC.

Table 5-11. Table of signal intensities of normal and tumour tissues in cDWI sequences.

cDWI	Tissue	Median signal intensity (IQR)	Tumour vs normal† (p-value)
Mean ADC ($\times 10^{-6}$ mm ² /sec)	Normal	1526 \pm 250*	<0.001
	Tumour	960 (8371068)	
Minimum ADC ($\times 10^{-6}$ mm ² /sec)	Tumour	784 \pm 208*	
Mean b1000 SI	Normal	147 (125189)	0.935
	Tumour	147 (128174)	
Mean b1500 SI	Normal	83.4 (68100)	<0.001
	Tumour	109 \pm 32*	

* Mean \pm Standard Deviation, † Wilcoxon signed rank test

Table 5-12. Table of signal intensities of normal and tumour tissues in sDWI sequences.

sDWI	Tissue	Median signal intensity (IQR)	Tumour vs normal* (p-value)
Mean ADC	Normal	1258 (10741387)	<0.001
	Tumour	766 (672887)	
Minimum ADC	Tumour	514 (374698)	
Mean b1000 SI	Normal	156 (138179)	<0.001
	Tumour	174 (154206)	
Mean b1500 SI	Normal	139 (125 160)	<0.001
	Tumour	120 (102 122)	

* Wilcoxon signed rank test

A selection of histograms of signal intensities from tumour and non-tumour tissue of different cDWI and sDWI sequences are demonstrated in Figure 5-27 A-D. Most distributions were non-parametric, but the signal intensity of tumour on b1500 images, the minimum tumour ADC, and the mean ADC value of non-tumour tissue on the cDWI sequence demonstrated a normal distribution.

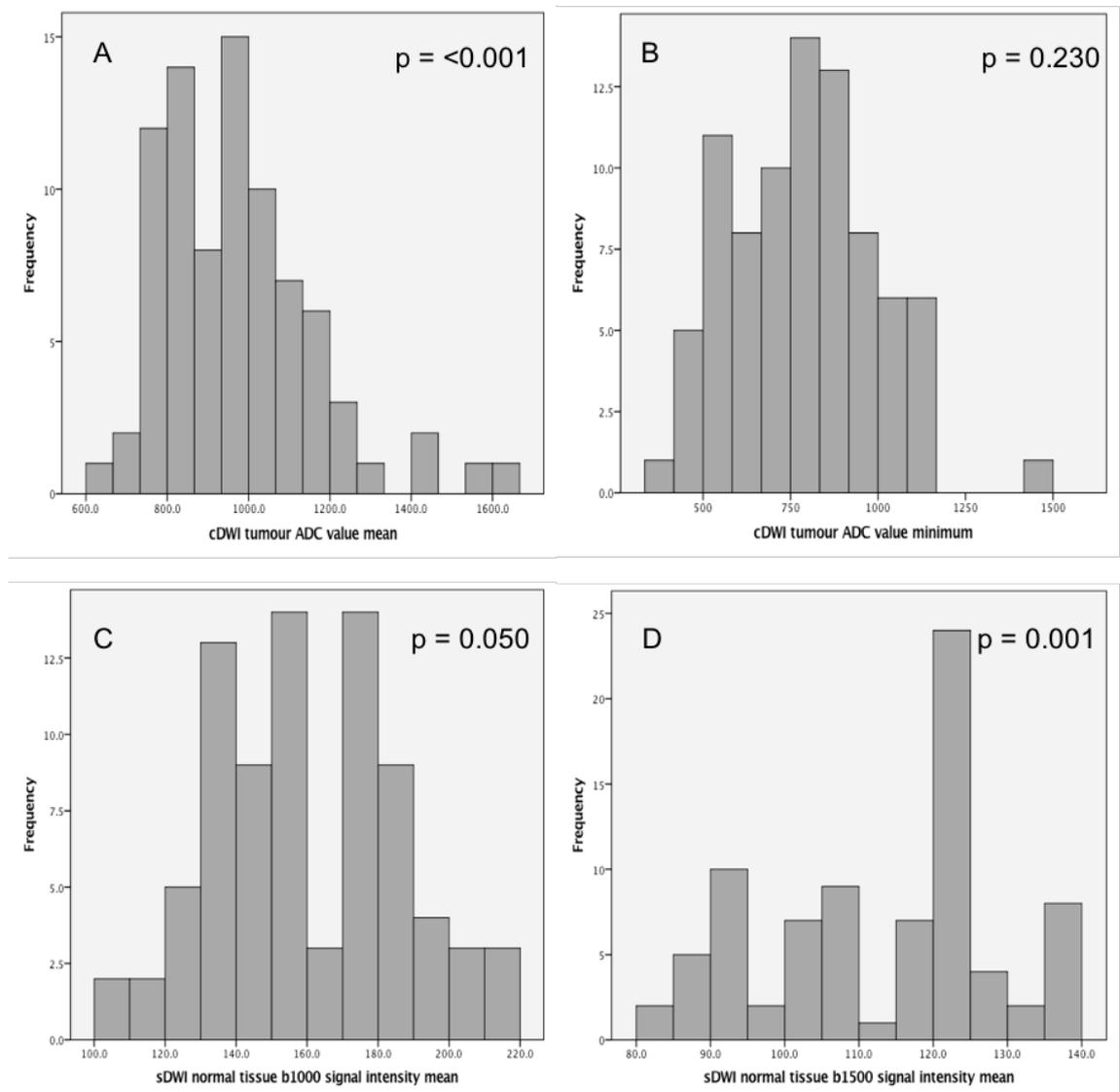


Figure 5-27. Histograms of selected signal intensities demonstrating parametric and non-parametric distributions and their Shapiro-Wilk test p-value. A - Mean ADC value of tumour, B - Minimum ADC value of tumour, C - b1000 and D – b1500 signal intensity of normal tissue.

Noise measurement

The measure of noise was the standard deviation of the SI of a ROI drawn over obturator internus. This was performed for the ADC map, b1000 and b1500 sequences of both cDWI and sDWI, and for which the results are demonstrated in Table 5-13 with example histograms demonstrated in Figure 5-28 A and B. There was more noise in the ADC maps of both sequences with lower noise levels at b1500 compared to b1000 images. There was significantly higher noise in the cDWI ADC maps compared to sDWI, but the opposite was observed in the b1000 and b1500 images (all comparisons $p = <0.001$, Wilcoxon Signed Rank test)

Table 5-13. Median noise measurements of sDWI and cDWI sequences.

	Sequence	Median noise level (IQR)
sDWI	ADC map	93.4 (81.8 103.5)
	b1000	9.2 (8.59 9)
	b1500	8.0 (7.48 4)
cDWI	ADC map	127.2 (112.3 165.6)
	b1000	5.7 (4.96 0)
	b1500	4.2 (3.35 3)

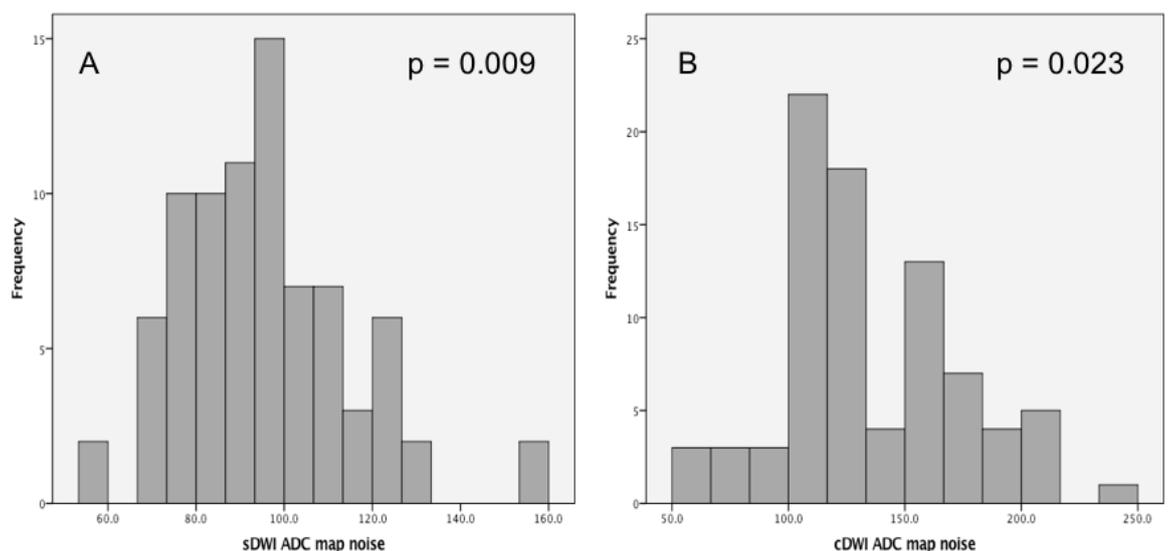


Figure 5-28. Histogram of noise level of A - sDWI ADC map and B – cDWI ADC maps. The p-value represents the Shapiro-Wilk test indicating a non-parametric distribution.

Tumour quantitative characteristics

The signal-to-noise ratio (SNR), contrast-to-noise (CNR), and contrast ratio (CR) of lesions on the ADC map, b1000 and b1500 sequences of both sDWI and cDWI are demonstrated in Table 5-14.

There were 83 lesions included in the cDWI group and 81 lesions in the sDWI for reasons described on page 164. Most of the tumour quantitative parameters were not normally distributed as suggested by the p-value of the Shapiro-Wilk test. A selection of histograms are demonstrated in Figure 5-29A-F illustrating examples of normal and not normal distribution.

Table 5-14. Tumour SNR, CNR and CR on cDWI and sDWI sequences and comparison between DWI sequences.

Sequence		cDWI	sDWI	Comparison of cDWI & sDWI
		Median (IQR)	Median (IQR)	Wilcoxon Signed Rank Test p-value
ADC map	SNR	7.4 (5.8–8.9)	8.8 ± 2.5*	<0.001
	CNR	-4.1 (-5.6–2.6)	-4.65 ± 2.82*	0.770
	CR	-0.22 ± 0.11*	-0.21 ± 0.13*	0.339†
b1000	SNR	25.7 (21.5–35.4)	18.9 (16.8–22.8)	<0.001
	CNR	-0.57 (-4.12–6.14)	2.22 (-0.21–3.81)	0.054
	CR	0.003 ± 0.15*	0.06 (0.01–0.10)	0.001
b1500	SNR	23.2 (18.0–31.8)	17.3 (15.3–19.8)	<0.001
	CNR	4.30 (-0.08–10.74)	3.08 (1.35–4.86)	0.190
	CR	0.12 ± 0.24*	0.11 ± 0.08*	0.151†

*mean ± standard deviation, †paired samples t-test

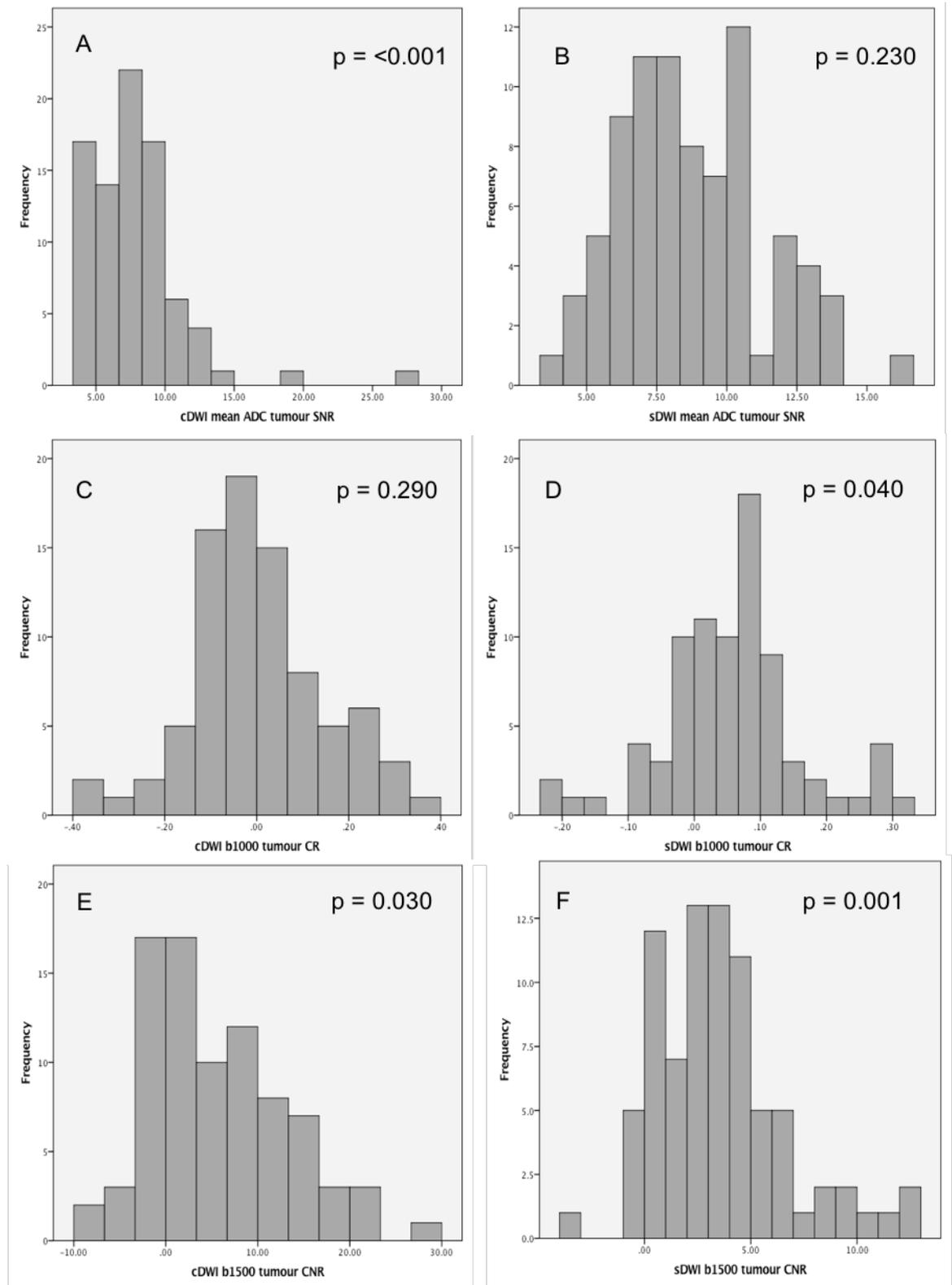


Figure 5-29. Histogram of selected tumour quantitative results of different sequences with Shapiro-Wilk test p-values demonstrated. A - cDWI ADC SNR, B - sDWI ADC SNR, C - cDWI b1000 CR, D - sDWI b1000 CR, E - cDWI b1500 CNR, F - sDWI b1500 CNR.

Comparison of quantitative results of sDWI and cDWI sequences

The cDWI and sDWI results were compared using the Wilcoxon-Signed Rank or paired t-tests. The SNR of the ADC map lesions was significantly higher on sDWI than cDWI, but the opposite was observed on the b1000 and b1500 images. The CR was significantly higher at b1000 with sDWI, but this difference was not observed at b1500. There was no significant difference between the CNR results between sDWI and cDWI on any sequence (Table 5-14).

The median mean ADC value of tumour was $766 \times 10^{-6} \text{ mm}^2/\text{sec}$ on sDWI ADC maps and $960 \times 10^{-6} \text{ mm}^2/\text{sec}$ on cDWI maps which was a significant difference ($p = <0.001$, Wilcoxon Signed Rank test) (Table 5-11 and Table 5-12, Figure 5-30). There was also a significant difference between the minimum ADC values of tumour between the two DWI sequences ($p = <0.001$, Wilcoxon Signed Rank test) (Figure 5-31).

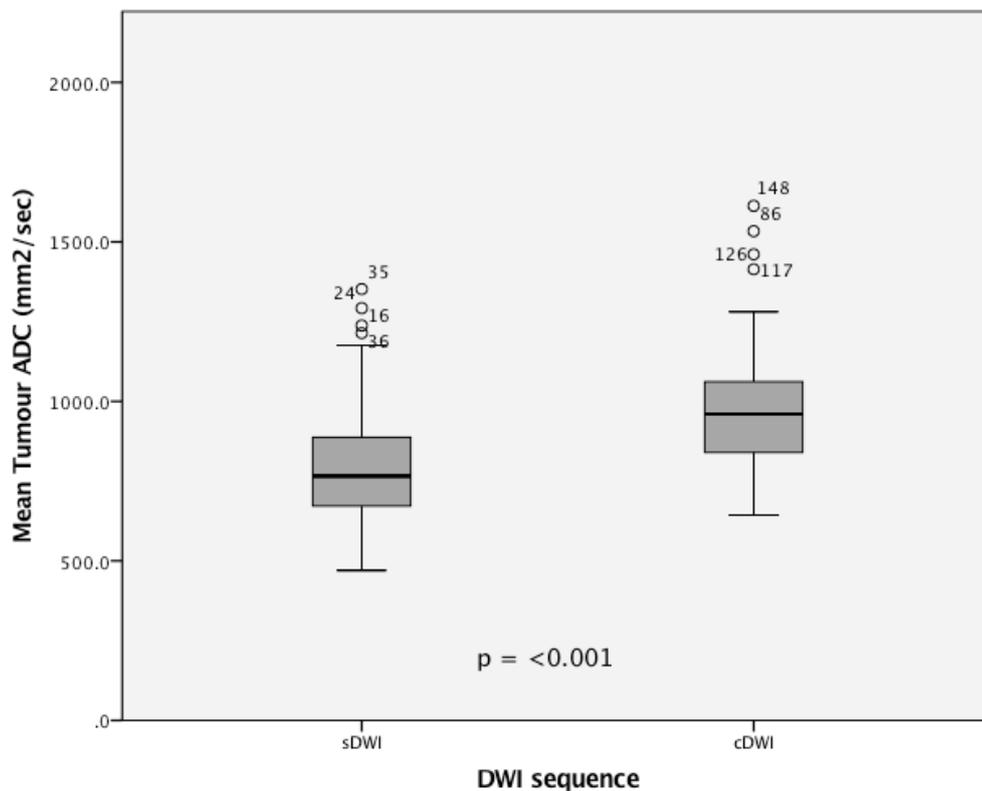


Figure 5-30. Boxplot of mean tumour ADC values of sDWI and cDWI ADC maps.

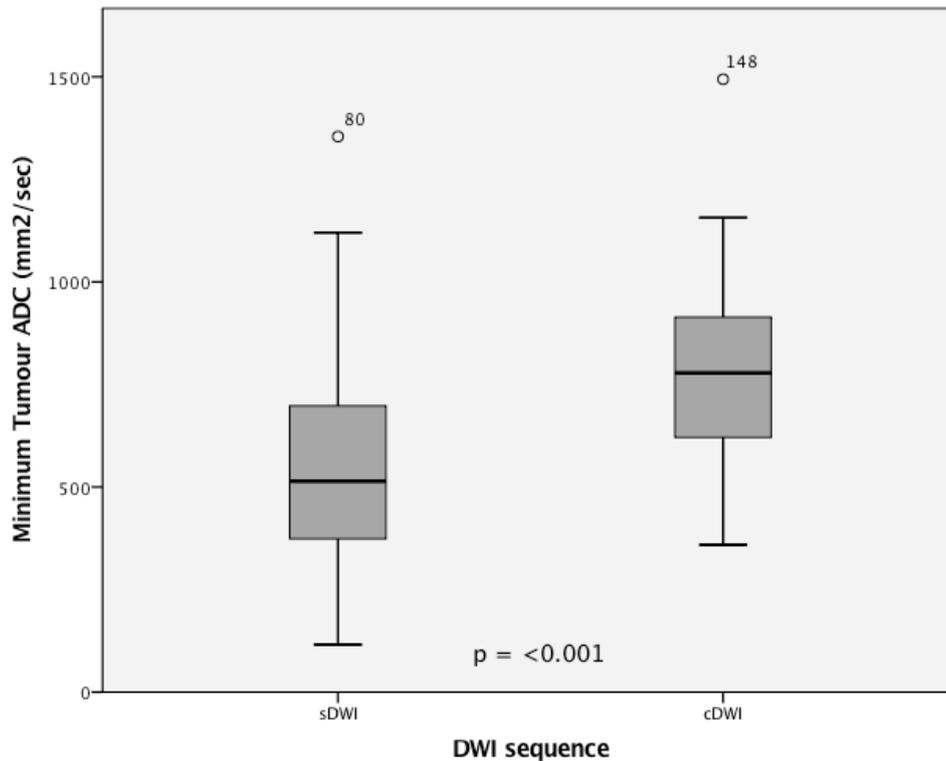


Figure 5-31. Boxplot of minimum tumour ADC values of sDWI and cDWI ADC maps.

Comparison of tumour quantitative results of b1000 and b1500 images

The SNR of lesions on b1000 images was significantly higher than on b1500 for both cDWI and sDWI (Figure 5-32 A and B, $p = <0.001$ – Wilcoxon signed rank). There was a significant improvement in tumour CNR (Figure 5-33 A and B) and CR (Figure 5-34 A and B) on b1500 images compared to b1000 for both cDWI and sDWI sequences ($p = <0.001$). The comparison of CR on cDWI images was performed using a paired samples t-test and the other comparisons of CNR and CR with a Wilcoxon Signed Rank test.

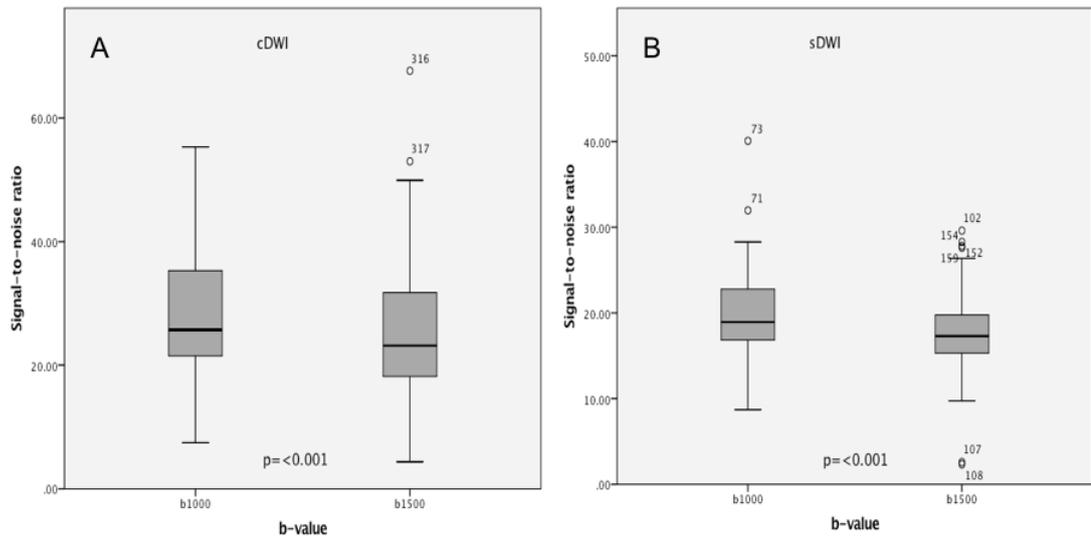


Figure 5-32. Comparison of SNR of tumour on b1000 and b1500 images using a Wilcoxon Signed Rank test. A - cDWI, B – sDWI.

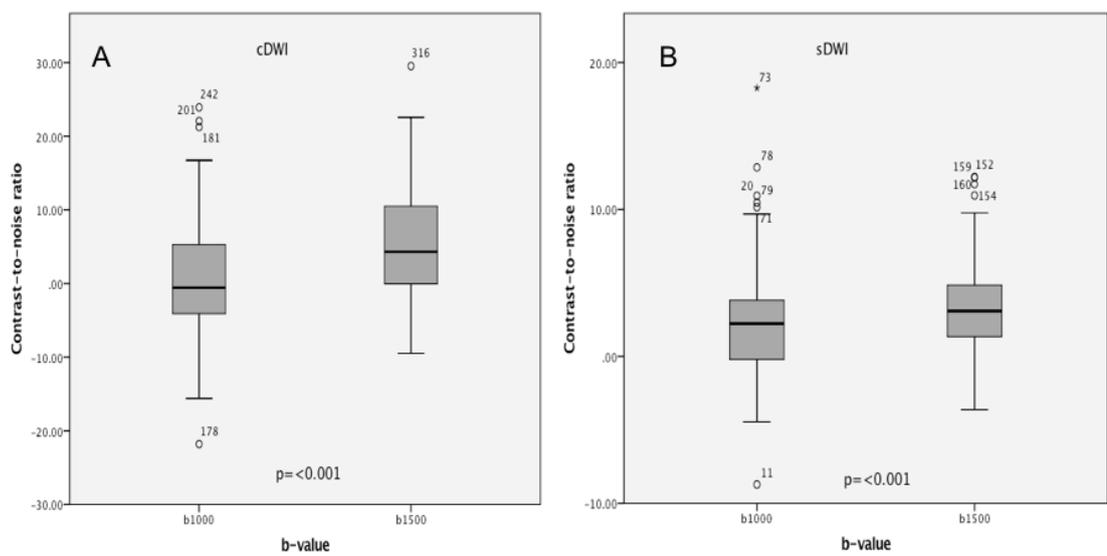


Figure 5-33. Comparison of CNR of tumour on b1000 and b1500 images using a Wilcoxon Signed Rank test. A - cDWI, B – sDWI.

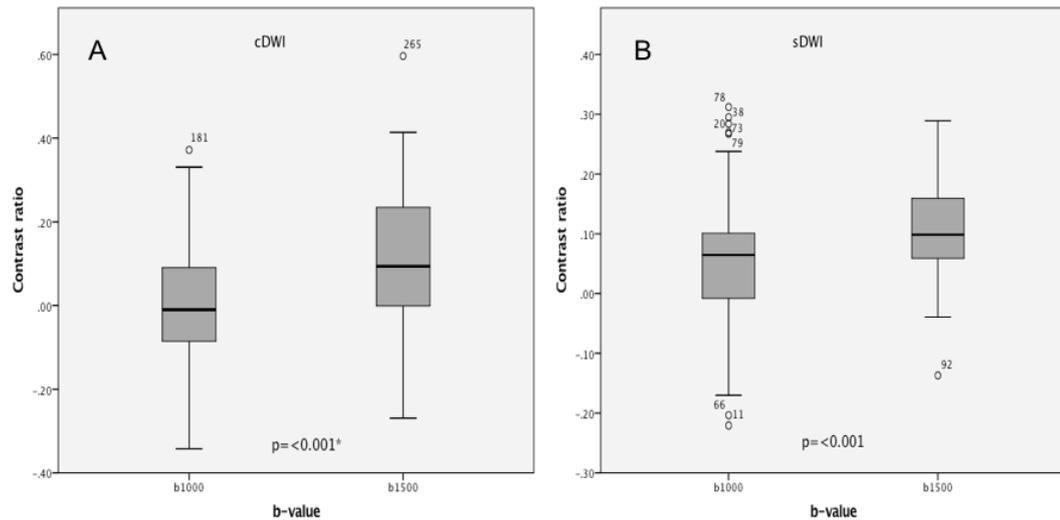


Figure 5-34. Comparison of CR of tumour on b1000 and b1500 images. A - cDWI (paired samples t-test), B - sDWI (Wilcoxon Signed Rank test).

Receiver Operating Characteristic curves of mean ADC values.

The ROC curves of the mean ADC value of cDWI and sDWI ADC map are demonstrated in Figure 5-35 and Figure 5-36, respectively. The area under the ROC curve (AUC) for the cDWI sequence was 0.96 (95% CI, 0.93–0.99) and for the sDWI sequence was 0.90 (0.84–0.96) which was not significantly different ($p = 0.07$, De Long's test for two correlated ROC curves). For the cDWI ADC map an ADC cut-off for tumour of less than $1153 \times 10^{-6} \text{ mm}^2/\text{sec}$ results in a sensitivity of 0.86 and specificity of 0.98. For the sDWI ADC map an ADC cut-off of for tumour of less than $898 \times 10^{-6} \text{ mm}^2/\text{sec}$ results in a sensitivity of 0.77 and a specificity of 0.93.

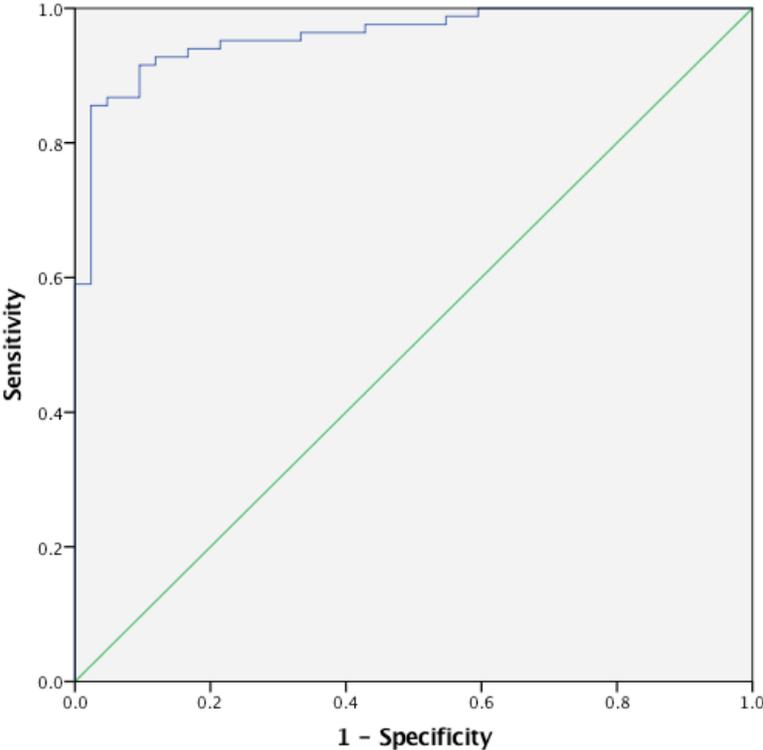


Figure 5-35. Receiver operating characteristic curve of mean ADC values of the cDWI sequence.

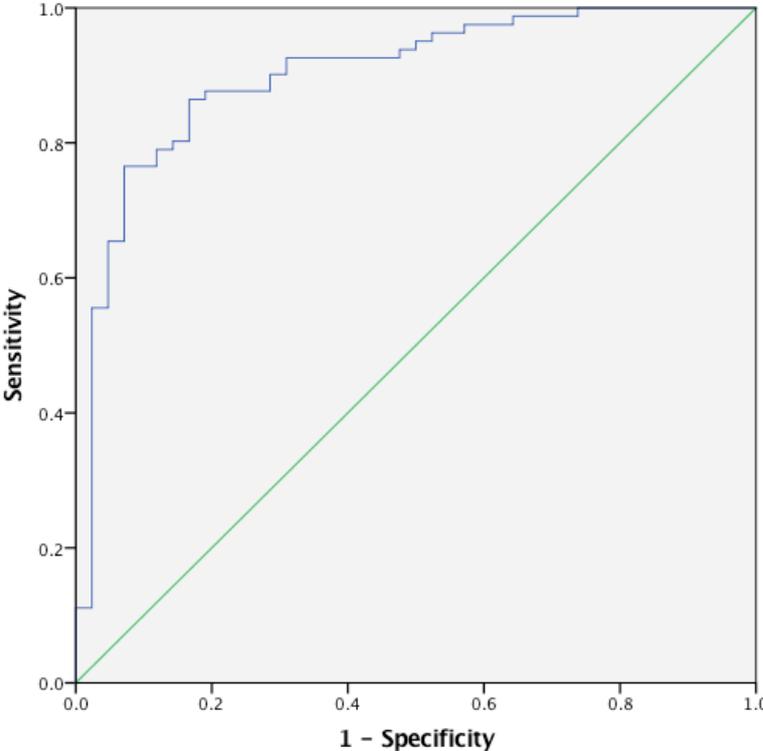


Figure 5-36. Receiver operating characteristic curve of mean ADC values of the sDWI sequence.

Quantitative results of tumour index lesions

The mean and minimum ADC values of index tumours were significantly different from non-index lesions, as was the SNR on ADC maps of both cDWI and sDWI (Table 5-15 and Table 5-16). The CNR of index lesions on sDWI ADC maps and CR of index lesions on cDWI ADC maps were significantly lower than non-index lesions. Signal intensity, SNR, CNR and CR of index lesions and non-index lesions were not significantly different on b1000 and b1500 images.

Table 5-15. Comparison quantitative imaging characteristics of index and non-index tumours on cDWI sequence.

cDWI		Index lesion Median (IQR)	Non-index lesion Median (IQR)	Difference p-value
ADC map	ADC mean ($\times 10^{-6}$ mm ² /sec)	852 (795–960)	1028 (946–1124)	<0.001†
	ADC min ($\times 10^{-6}$ mm ² /sec)	646 (546–772)	889 \pm 176*	<0.001†
	SNR	6.6 (4.9–7.8)	8.4 (6.7–9.6)	0.003†
	CNR	-4.91 \pm 2.94*	-3.85 \pm 2.49*	0.080‡
	CR	-0.26 \pm 0.11*	-0.21 \pm 0.09*	0.002‡
b1000	SI	162 \pm 46*	144 (126–157)	0.189†
	SNR	26.6 (20.9–37.7)	27.0 \pm 8.2*	0.444†
	CNR	2.36 \pm 9.00*	-0.70 \pm 7.25*	0.103‡
	CR	0.03 \pm 0.16*	-0.01 (-0.10–0.04)	0.282†
b1500	SI	114 \pm 36*	99 (83–116)	0.111†
	SNR	23.1 (19.6–30.6)	24.6 \pm 10.0*	0.535†
	CNR	6.80 \pm 7.94*	3.22 \pm 8.72*	0.194‡
	CR	0.14 \pm 0.28*	0.10 \pm 0.16*	0.191‡

* Mean \pm Standard deviation, † Mann-Whitney U test, ‡ Independent samples t-test

Table 5-16. Comparison quantitative imaging characteristics of index and non-index tumours on sDWI sequence.

sDWI		Index lesion	Non-index lesion	Difference
		Median (IQR)	Median (IQR)	p-value
ADC map	ADC mean ($\times 10^{-6}$ mm ² /sec)	724 \pm 141*	885 \pm 199*	<0.001‡
	ADC min ($\times 10^{-6}$ mm ² /sec)	426 (340–517)	623 \pm 238*	<0.001†
	SNR	7.9 \pm 2.3*	9.6 \pm 2.5*	0.003‡
	CNR	-5.21 \pm 2.73*	-4.11 \pm 2.84*	0.014‡
	CR	-0.24 \pm 0.11*	-0.18 \pm 0.13*	0.080‡
b1000	SI	185 \pm 43*	171 (154–206)	0.487†
	SNR	20.1 \pm 5.0*	18.9 (16.9–22.2)	0.613†
	CNR	2.78 \pm 3.70*	1.49 (-.60–3.09)	0.236†
	CR	0.07 \pm 0.09*	0.04 (0.01–0.11)	0.218†
b1500	SI	145 \pm 32*	138 \pm 31*	0.386‡
	SNR	17.8 (16.0–20.2)	17.1 \pm 4.6*	0.375†
	CNR	3.78 (1.38–5.87)	3.06 \pm 3.01*	0.183†
	CR	0.11 (0.06–0.18)	0.08 (0.08)	0.142†

* Mean \pm Standard deviation, † Mann-Whitney U test, ‡ Independent samples t-test

The AUC of the sDWI sequence was 0.96 (95%CI, 0.91–1.00) and for cDWI was 0.97 (95%CI, 0.94–1.00) which was not significantly different ($p=0.56$). For index tumours, an ADC cut-off of 789 $\times 10^{-6}$ mm²/sec gives a sensitivity of 0.75 and a specificity of 0.97 for sDWI, and a cut-off of 1122 $\times 10^{-6}$ mm²/sec gives a sensitivity and specificity of 0.90 and 0.97 for cDWI, respectively (Figure 5-37).

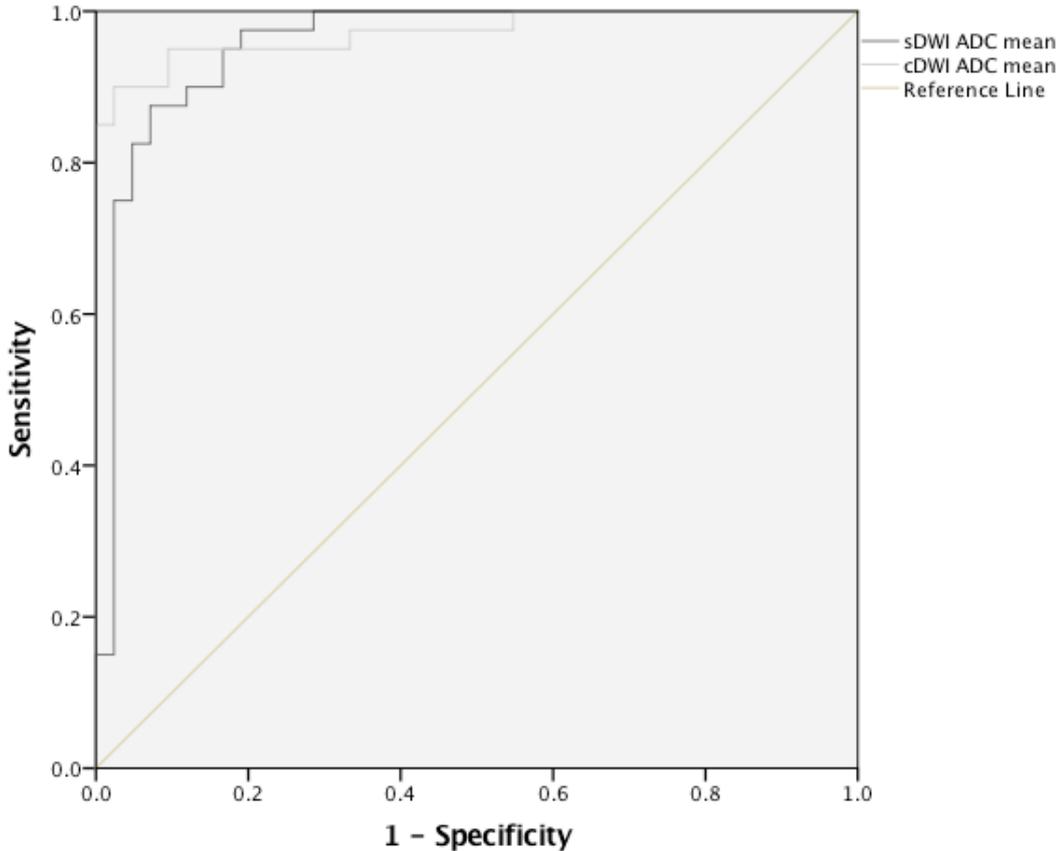


Figure 5-37. ROC curve of mean ADC value of index lesions on each ADC map.

Influence of haemorrhage on tumours

The mean tumour ADC was significantly higher in lesions within haemorrhage when compared to those outwith, and this significance is observed in both cDWI and sDWI ADC maps (Table 5-17 and Table 5-18). This significant difference is also demonstrated in the minimum ADC value of tumours on cDWI. Significantly increased contrast (both CNR and CR) was observed in lesions not affected by haemorrhage on the sDWI ADC map, when compared to tumours affected by haemorrhage. There was no significant effect of haemorrhage on lesions on the b1000 or b1500 images on cDWI or sDWI. Wilcoxon signed rank test comparing sDWI and cDWI mean and minimum tumour ADC values in tumours without evidence of haemorrhage was statistically significant ($p < 0.001$).

Table 5-17. Quantitative imaging parameters and comparison of tumours situated within and outwith haemorrhage on cDWI sequence.

cDWI		No haemorrhage Median (IQR)	Haemorrhage present Median (IQR)	Difference p-value
ADC map	ADC mean ($\times 10^{-6}$ mm ² /sec)	918 (816–1037)	1012 (959–1137)	0.011†
	ADC min ($\times 10^{-6}$ mm ² /sec)	750 \pm 195*	886 \pm 216*	0.009‡
	SNR	7.2 (5.7–8.9)	8.2 (6.3–9.4)	0.209†
	CNR	-4.42 (-5.58–-2.69)	-3.89 \pm 3.26*	0.414†
	CR	-0.23 \pm 0.10*	-0.19 \pm 0.13*	0.132‡
b1000	SI	155 \pm 47*	148 (134–183)	0.660†
	SNR	28.5 \pm 11.1*	25.05 (22.5–35.2)	0.834†
	CNR	-0.41 (-4.066.32)	-0.63 \pm 7.30*	0.444†
	CR	-0.01 \pm 0.16*	-0.01 \pm 0.12*	0.579‡
b1500	SI	109 \pm 34*	106 \pm 24*	0.814‡
	SNR	23.2 (18.0–33.0)	25.0 \pm 10.2*	0.900†
	CNR	5.96 \pm 7.89*	4.87 \pm 6.37*	0.568‡
	CR	0.12 \pm 0.15*	0.12 \pm 0.17*	0.992‡

* Mean \pm Standard deviation, † Mann-Whitney U test, ‡ Independent samples t-test

Table 5-18. Quantitative imaging parameters and comparison of tumours situated within and outwith areas of haemorrhage on sDWI sequence.

sDWI		No haemorrhage Median (IQR)	Haemorrhage present Median (IQR)	Difference p-value
ADC map	ADC mean ($\times 10^{-6}$ mm ² /sec)	741 (649–871)	909 \pm 197*	0.005†
	ADC min ($\times 10^{-6}$ mm ² /sec)	498 (371–637)	621 \pm 245*	0.071†
	SNR	8.6 \pm 2.6*	9.3 \pm 2.2*	0.285‡
	CNR	-5.14 \pm 2.54*	-3.16 \pm 3.16*	0.002‡
	CR	-0.23 \pm 0.11*	-0.13 \pm 0.14*	0.006‡
b1000	SI	176 (155–206)	177 \pm 41*	0.718†
	SNR	19.1 (16.8–22.2)	19.5 \pm 4.92*	0.793†
	CNR	2.50 (-0.48–4.27)	2.17 \pm 3.44*	0.801†
	CR	0.06 \pm 0.11*	0.05 \pm 0.10*	0.882‡
b1500	SI	139 (125–161)	139 \pm 28*	1.000†
	SNR	17.29 (15.8–19.8)	17.14 \pm 4.12*	0.870†
	CNR	3.09 (1.27–4.86)	3.41 \pm 2.55*	0.987†
	CR	0.11 \pm 0.09*	0.09 (0.07–0.14)	0.784†

* Mean \pm Standard deviation, † Mann-Whitney U test, ‡ Independent samples t-test

The ROC curve demonstrated an AUC of 0.92 (95%CI, 0.87–0.98) and 0.97 (95% CI, 0.95–1.00) for lesions not affected by haemorrhage for sDWI and cDWI sequences respectively (Figure 5-38). An ADC cut-off on the sDWI sequence of less than 894×10^{-6} mm²/sec for tumour and non-tumour tissue results in a sensitivity of 0.83 and specificity of 0.93. For cDWI, an ADC cut-off of less than 1153×10^{-6} mm²/sec gives a sensitivity of 0.89 and specificity of 0.98.

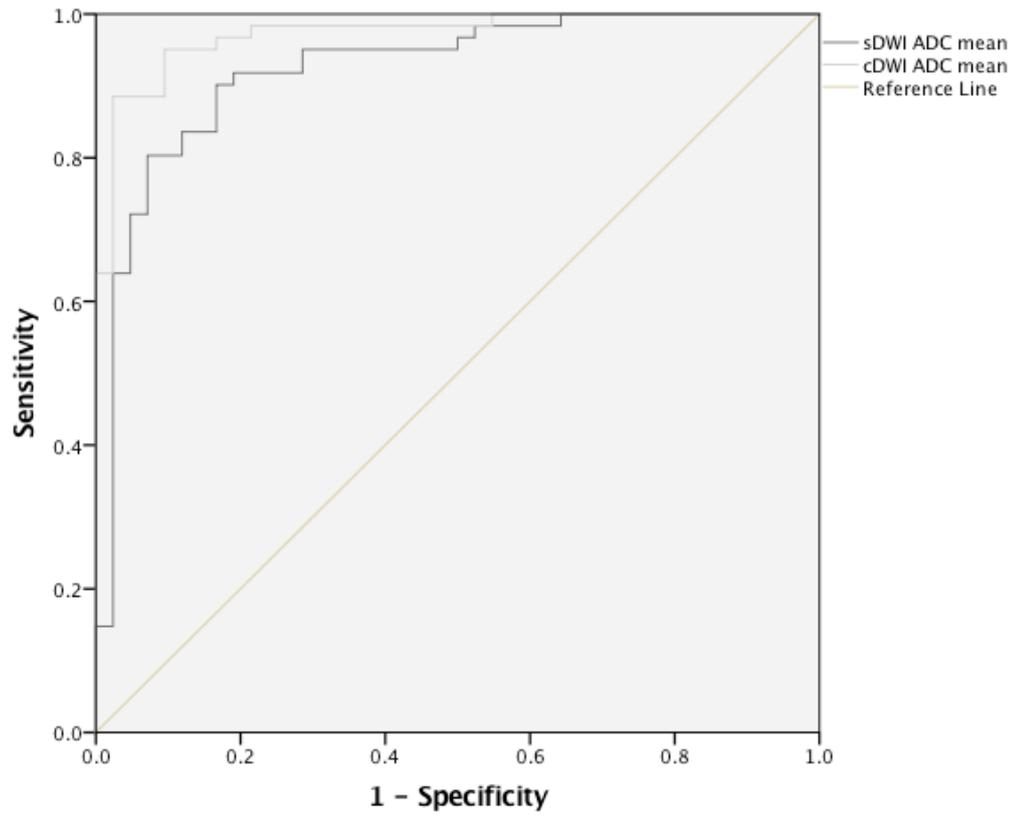


Figure 5-38. ROC curve of mean ADC value of sDWI and cDWI sequences including only tumours not affected by haemorrhage.

Correlation of tumour ADC value and tumour aggressiveness

There is a trend towards reducing mean and minimum ADC values as the tumour grade rises (Table 5-19, Figure 5-39 - Figure 5-42), but no Spearman correlation coefficient reached statistical significance. The strongest correlation was a ρ of -0.210 ($p = 0.057$) when the minimum ADC value of tumours on the cDWI sequence were correlated with Gleason Grade group (Figure 5-39B). The correlation coefficient did not improve when the Gleason Grade Groups were split into Group 2 lesions and those more aggressive.

Table 5-19. Mean and minimum ADC values of tumours of different Gleason Grade Group and for tumours of Gleason Grade 3+4 and those above.

		cDWI		sDWI	
	Number of tumours	Mean ADC Mean \pm SD ($\times 10^{-6}$ mm ² /sec)	Minimum ADC Mean \pm SD ($\times 10^{-6}$ mm ² /sec)	Mean ADC Mean \pm SD ($\times 10^{-6}$ mm ² /sec)	Minimum ADC Mean \pm SD ($\times 10^{-6}$ mm ² /sec)
Normal tissue		1526 \pm 250		1258 (1074–1387)*	
Grade Group					
2	57	979(837–1088)*	805 \pm 214	824 \pm 197	567 \pm 250
3	18	929 \pm 187	740 \pm 202	777 \pm 180	506 \pm 198
4	4	961 (864–965)*	720 \pm 156	695 \pm 148	407 \pm 224
5	2	885 \pm 119	571 \pm 16	776 \pm 72	428 \pm 10
Gleason Score					
3 + 4	57	979 (837–1088)*	805 \pm 214	824 \pm 256	567 \pm 333
> 3 + 4	26	891 (823–970)*	722 \pm 189	763 \pm 209	483 \pm 218

SD – Standard deviation * Median (IQR) as data is non-parametric.

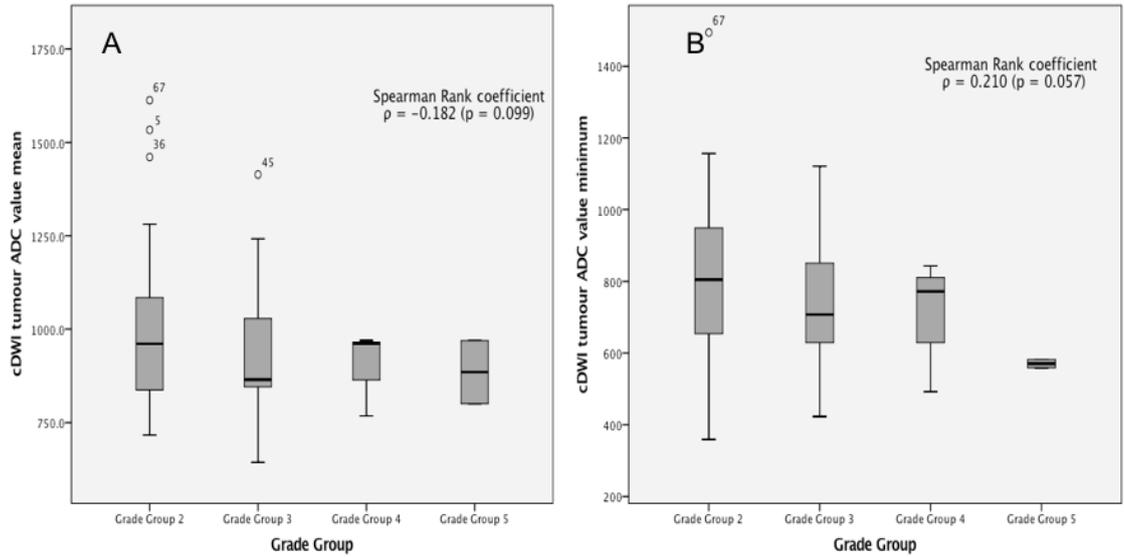


Figure 5-39. Boxplot and Spearman Rank correlation coefficient of ADC of different tumour grades on cDWI. A - Mean tumour ADC, B - Minimum tumour ADC.

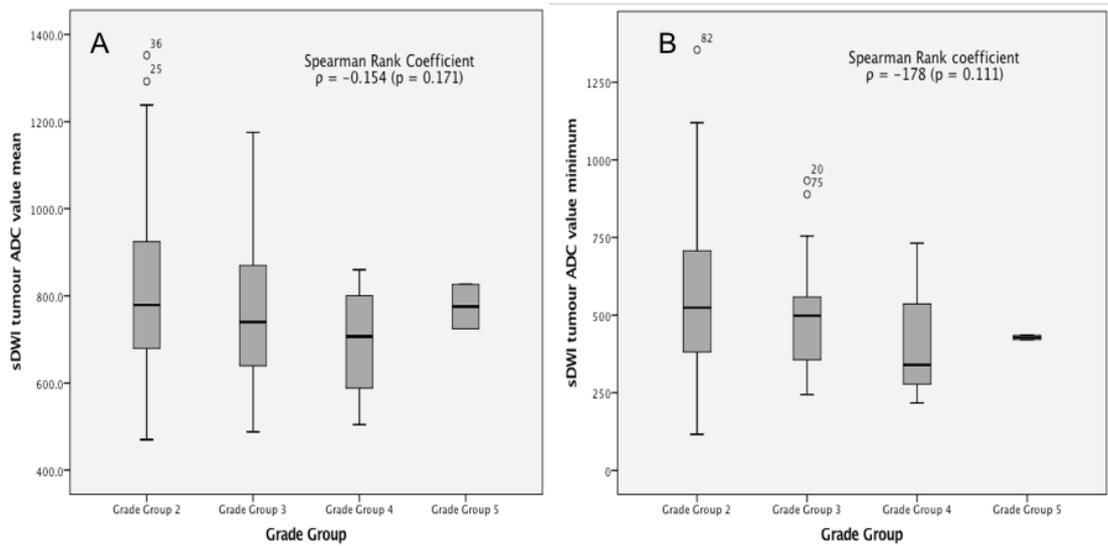


Figure 5-40. Boxplot and Spearman Rank correlation coefficient of ADC of different tumour grades on sDWI. A - Mean tumour ADC, B - Minimum tumour ADC.

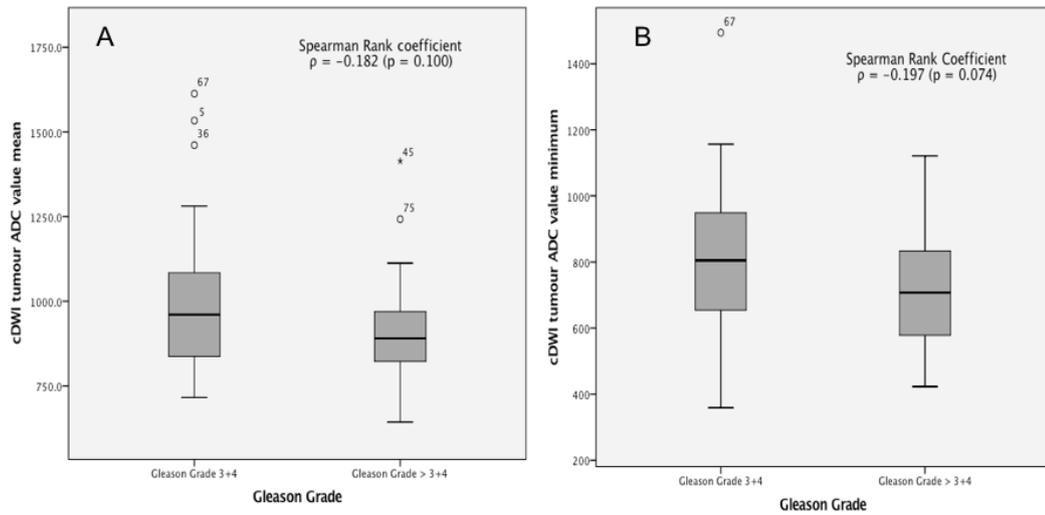


Figure 5-41. Boxplot and Spearman Rank correlation coefficient of tumour ADC of Gleason Score 3+4=7 lesions and those greater than Gleason Score 3+4=7 on cDWI. A - Mean tumour ADC, B - Minimum tumour ADC.

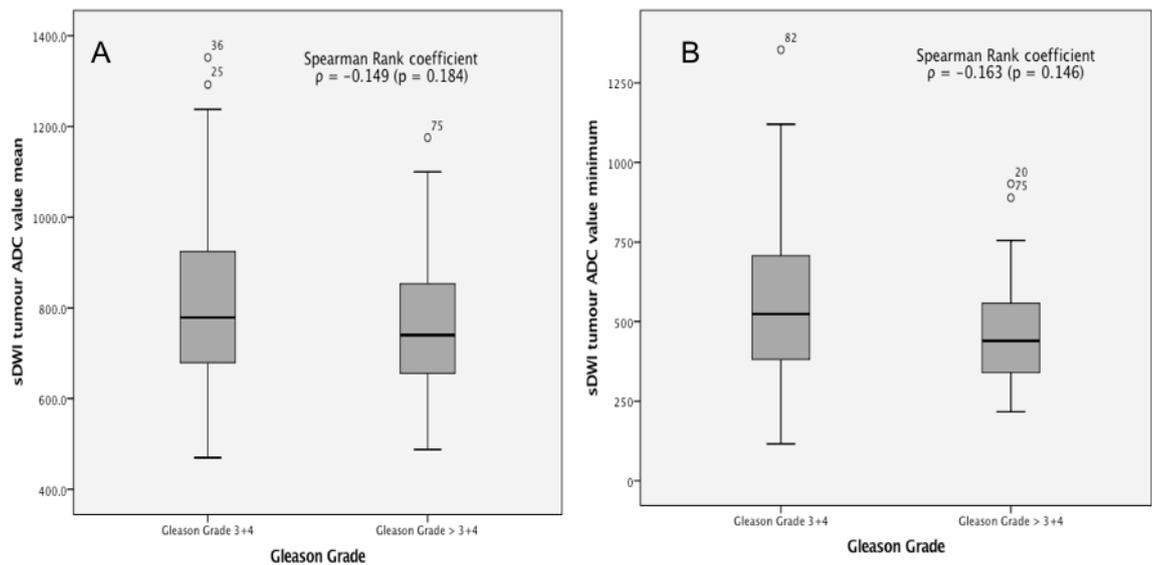


Figure 5-42. Boxplot and Spearman Rank correlation coefficient of tumour ADC of Gleason Score 3+4=7 lesions and those greater than Gleason Score 3+4=7 on sDWI. A - Mean tumour ADC, B - Minimum tumour ADC.

Chapter 6 Discussion

Accuracy of the DWI sequences

Lesion-based accuracy

Regardless of reader, addition of T2WI sequence, or whether a 'stringent' or 'clinical' lesion accuracy assessment was conducted, more index lesions were identified by conventional DWI than small FOV DWI. The differences between the comparable DWI sequence (ie the same reader, with or without T2WI) were often small. For example, 31 index lesions were seen by reader 2 with cDWI and 30 with sDWI when subject to a 'stringent' analysis. However, when a 'clinical' assessment was performed the difference was 3 tumours (34 versus 31 index lesions, respectively). In a clinical setting the difference between missing three index lesions for every 40 patients using small FOV DWI compared to conventional DWI could be considered significant, particularly as these are the tumours which are largest, most advanced or most aggressive. Regardless of which DWI sequence was used, about one-fifth of index lesions were not correctly identified and localised within the prostate, with more identified, but perhaps not accurately localised, within the hemi-gland. The addition of the T2WI sequence facilitated the identification of more index lesions for both DWI sequences and this result was observed in both readers.

There results of the diagnostic performance of both readers were similar when all lesions were assessed. They both found conventional DWI to be more sensitive; small FOV DWI to have a greater positive predictive value; and the 'clinical' assessment of the MRI reads was better than the 'stringent'. In general, the addition of the T2WI sequence improved sensitivity and PPV, with the exception that reader 1's cDWI 'clinical' and 'stringent' sensitivity assessment, and reader 2's sensitivity and PPV results changed very little. There was no significant difference between the results of the comparable cDWI and sDWI lesion accuracy analyses. The strongest evidence (p-

value 0.05–0.10) of a difference between DWI sequences was the higher PPV of reader 1's comparison of DWI sequences without T2WI, and reader 2's comparison of DWI sequences with T2WI in which small FOV out-performed the conventional DWI.

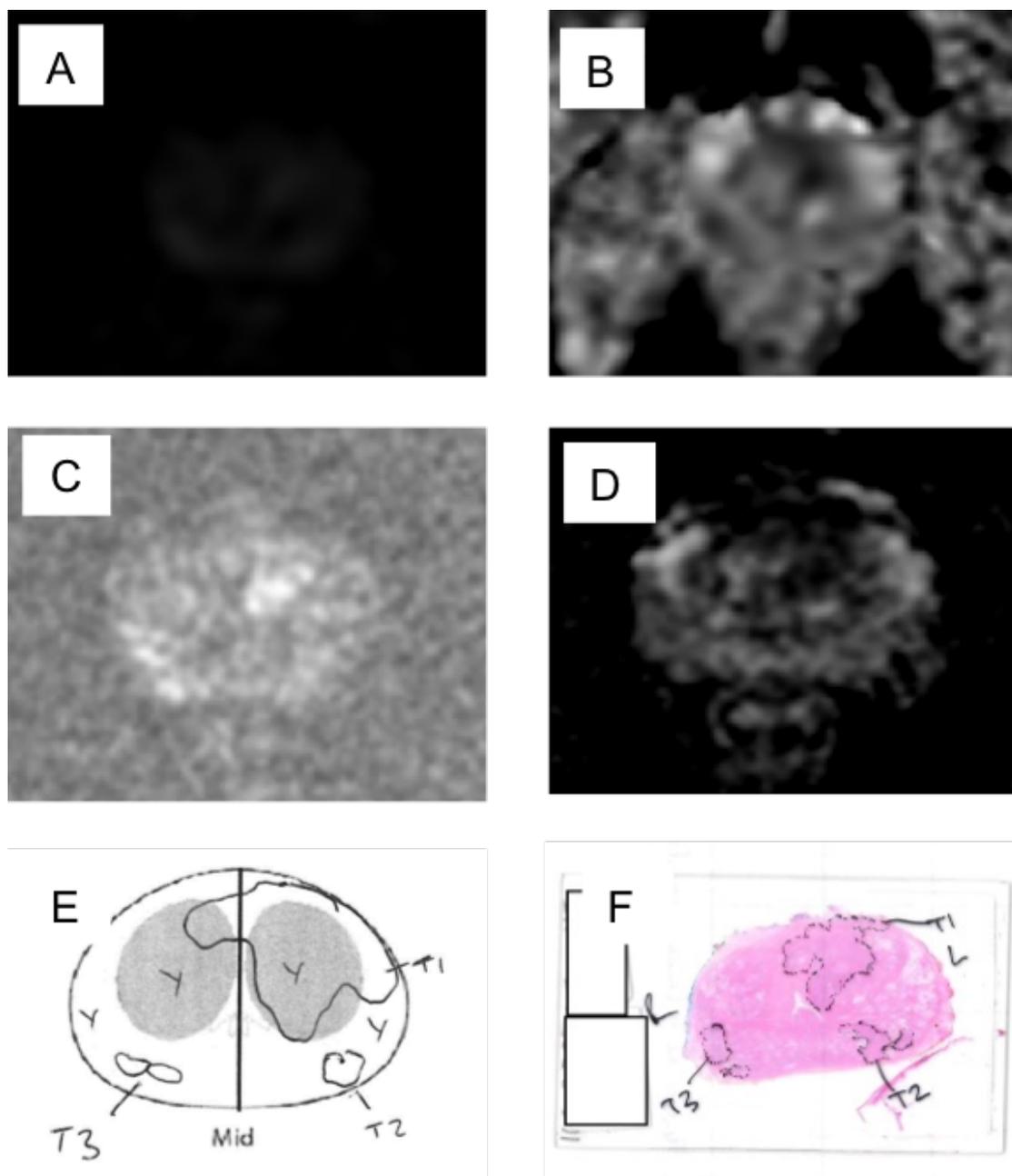


Figure 6-1. Multifocal prostate cancer. A - cDWI b1500, B - cDWI ADC map, C - sDWI b1500, D - sDWI ADCmap, E - Pathology template report, F - Image of pathology slide.

Figure 6-1 demonstrates a multifocal prostate cancer with 3 separate tumours in the mid-gland. The index lesion was the larger and identified by both readers on both DWI sequences, where as the 2nd lesion (T2 on Figure 6.1F) was a Gleason 3+4 tumour and seen by neither radiologist on

both DWI sequences. The third lesion (T3) was seen by both radiologists, but only on the conventional DWI sequence. This difference in tumour identification between DWI sequences is potentially because of the apparent low ADC signal in the right peripheral zone on cDWI with no conspicuous lesion on sDWI. Visually, the ADC map has been described as being more accurate than the trace DWI images [81] and therefore the more conspicuous lesion on cDWI ADC maps perhaps has resulted in the improved lesion detection and sensitivity.

The greatest hindrance to determining whether the differences in sensitivity and PPV between the DWI sequences was significant was the small number of lesions, which is evident in the wide confidence intervals in the results. Attempts to maximise the number of included lesions was made initially by instructing the histopathologists to assess for multifocality, rather than just the index lesion, and having a size cut-off for inclusion of at least 5mm in the maximum axial diameter. This is in comparison to some other studies which also have a volume cut-off of 0.5cm^3 [202] and therefore excluded smaller tumours. Some studies which assess for index lesions alone and not for multiple lesions calculate and cite the PPV [81,134]. In the FODIP study, the PPV for index lesion detection was not calculated as the readers were instructed at outset to assess the prostate for multiple lesions rather than the single index tumour. This would allow the overcalling (false positive) of more than one lesion. Whereas if specifically assessing just for a single index lesion an incorrectly located lesion becomes both a false positive and false negative.

Overall, despite the small number of lesions included, the similarity of the results between readers, regardless of presence of T2WI, adds to the confidence of the findings indicating the conventional DWI sequence is better at lesion detection, but has a greater likelihood of overcalling tumours.

Likert scoring system

Both readers were more definite in scoring the presence or absence of tumour, with more Likert 1 and 5 scores assigned to sectors, compared to Likert 2 and 4 lesions. The addition of the T2WI sequence did not alter the distribution of scores of reader 1, but had more of an effect on reader 2 with a reduction in Likert 1 lesions and an increase in Likert 5 lesions when T2WI was added. The lack of change is potentially due to the immediate reads of Protocols B and D following the read of Protocol A and C, respectively, for each case and will be addressed later in the discussion.

One clinically useful finding is the reduction in Likert 3 lesions when the T2WI sequence was added. Likert 3 lesions are indeterminate and are a diagnostic dilemma if tumour has not already been proven histologically. If a targeted prostate biopsy follows the report of an indeterminate lesion and returns negative for tumour, the urologist and patient have a difficult decision to make regarding re-biopsing the prostate, which is invasive and not without risk, or monitoring the patient with other tools such as serum PSA levels and PSA density.

The interplay between different Likert score cut-offs on the values of the 2 x 2 table for both readers behaved as expected. When a Likert 3 cut-off was used, with scores of 3 and above deemed tumour and less than 3 negative for tumour, this resulted in a greater number of true positive and false positive lesions, and greater sensitivity and smaller specificity. When the cut-off is changed to 4 the number of true and false positives fall, but false and true negatives rise, resulting in a fall in sensitivity and a rise in specificity when compared with a cut-off score of 3.

The interplay of Likert score cut-offs is important because in some instances it may be of more use to have a more specific or sensitive test. For example, in prostate cancer management if the MRI of the prostate was performed to stage biopsy proven prostate cancer it may be of more use to have a more sensitive test with fewer false negatives and a greater ability to identify, localise and accurately stage the known cancer. Conversely, in a pre-biopsy setting a more specific test

may be more beneficial where the ability to rule out clinically significant cancer is important to reassure the patient and avoid a biopsy. Likert scoring systems can be tailored to each patient dependant on their wishes. In the management of prostate cancer a Likert scoring system is regularly used when following the guidelines outlined in PI-RADS version 2 [67]. A grey area is frequently the management of Likert 3 lesions and much of the current research is focused on these patients and lesions. As the results demonstrate higher specificity than sensitivity, regardless of protocol or reader, it was decided for the lesion-based assessment a Likert 3 lesion and above would be considered positive for tumour to improve sensitivity.

There is variation in the use of Likert score cut-offs and the reporting of them amongst diagnostic test accuracy studies in prostate cancer. A review of most of the scoring systems used in the T2WI and DWI meta-analysis (Chapter 3, page 82) has shown that most used the same scoring system used in FODIP study, but some studies displayed the diagnostic performance of the tested sequences but did not state the cut-off used [83,104,187]. Only one study stated diagnostic performance of different cut-offs [180], some used a cut-off of 4 and above for tumour [85,177] and most used a cut-off of 3 [86,138,140].

Sector-based accuracy

The results of the FODIP study demonstrated an improvement in sensitivity using the conventional DWI sequence when compared to the small FOV DWI, which reached statistical significance with reader 1 (example demonstrated in Figure 6-1, Page 192). Conversely, specificity was higher with sDWI than cDWI and again reached statistical significance with reader 1. Although, there was higher PPV with small FOV DWI and NPV with conventional DWI, these differences did not reach statistical significance regardless of reader or presence of T2WI. There was a significant improvement in sensitivity with conventional DWI and reduction in specificity with sDWI for reader 2. Reader 1 found a significant improvement in sensitivity with the addition of T2WI to sDWI. Many of these significant differences persisted when Bonferroni correction for

multiple comparisons was applied to the McNemar tests. The AUC of the ROC curve analysis indicated the cDWI and sDWI sequences were fair diagnostic tests [153] with trends suggesting cDWI is a better diagnostic test than sDWI, but this did not reach statistical significance. Reader 1 found the AUC to significantly increase with the addition of T2WI to sDWI.

The results of the sectoral diagnostic performance indicate that conventional DWI is better at tumour detection, despite no significant difference between DWI in the quantitative tumour, except a higher sDWI contrast ratio on the b1000 images. The conventional DWI sequence demonstrated significantly higher tumour SNR on the DWI images, but lower SNR on the ADC maps. This is likely related to the higher levels of noise found in the ADC map of the conventional DWI sequence, but lower noise in the b1000 and b1500 sequences when compared to the sDWI sequence. There have been few studies directly comparing the qualitative diagnostic performance of the DWI images compared to the ADC maps [81,82]. In Rosenkrantz's diagnostic test accuracy study of high b-value DWI compared to high b-value ADC maps found that sensitivity improved with the b2000 image set compared to the ADC map, and there were less false positive lesions on the DWI maps [81]. In an earlier study by Rosenkrantz et al. [82] tumour contrast ratios were higher on ADC maps than on the b500 and b1000 sequences. If indeed high b-value images are the most sensitive then perhaps the significantly greater SNR of the conventional b1000 and b1500 images has led to the improved detection of tumours.

A further possible explanation for the difference in diagnostic performance of small FOV and conventional DWI is the use of high b-values which is associated with lower signal. As the b-value rises the signal in normal tissue falls more than the signal in cancerous tissue. However, for smaller fields-of-view the lower signal at higher b-values is compounded by an increase in the image noise. It is possible that at high b-values the signal is sufficiently low with a small FOV that it compromises the detection of tumour. This is further supported by the improved CNR of lesions with cDWI at b1500, but with sDWI at b1000. This could be further evaluated by repeating the

qualitative and quantitative accuracy study with ADC maps derived from the b100 and b1000 images for both small and large fields of view.

Challenges of the radiologic-histopathologic correlation

The precision of the sectoral and lesion-based accuracy results is dependent on the accuracy of the histopathological assessment and thereafter the correlation of the radiology and pathology findings. The validity of the Gleason Grades of each tumour and the size of the lesion is improved by the review by histopathologists, however, there were other points in the process in which error could have been introduced.

Firstly, the thickness of the prostates slices ranged between 4.5 and 7mm, and between these slices tumours may be sited and have been missed. In comparison, the slice thickness of the DWI and T2WI sequences was 3.6mm with no gaps. Therefore, MRI samples all the gland and histopathology risks missing lesions due to sampling of a small volume of the prostate with thicker slice. This could result in some of the lesions called false positive sectors incorrectly and were in fact true positives.

The variation in slice thickness is due to the technique of cutting the prostate which involves the free-hand slicing of the prostate in the axial plane after the apex has been cut-off. The freehand cutting technique has a greater chance of producing axial slices which are not perpendicular to the cranio-caudal axis of the prostate and do not quite correlate to the axial slices of the MRI. This could also lead to a mis-registration of the exact location of the tumour.

Determination of the location within the gland with respect to height and placement within the base, mid-gland and apex is a challenge for both MRI and histopathology assessment. There are no reliable anatomical landmarks at either assessment that differentiate between them. For both radiologists and histopathologists this was done by assessing how many slices the prostate has been divided into and then splitting the slices into the three groups. Although, histopathologists

traditionally consider the apex the slice which is separated and cut longitudinally. The potential for misplacement of lesions into different locations in the gland has implications for the accuracy of the assessment.

The correlation of the histopathology and radiology templates was also a challenge. In particular, the lesion analysis was difficult when there was more than one tumour in the same region of gland. The correlation was performed by KG and after the assessment of a few cases it led to the creation of some rules to increase the consistency of the correlation process. For example, if two tumours were in the same sector at histopathology for both to be seen at radiology there would have to be two tumours drawn. Also, if a tumour is described as two separate lesions by a radiologist, but one by a pathologist then the correlation was described as a true positive with no additional false positive. Finally, if a tumour was considered as one tumour by radiology, but two by histopathology then if both lesions called by the radiologist were in the same area the histopathologist lesion was drawn, this was considered a single true positive and no false positive.

Any discrepancies were put to the histopathology team and resolved in consensus (KG and HP2). For example, in one case a tumour was seen by both radiologists in the apex of the gland and reported as a Likert 5 lesions in all 4 protocols, but no tumour was reported in this region by histopathology. A request was then made to histopathology to re-evaluate the apex and a Gleason 4+3 lesion was seen on the re-assessment, which was the index tumour. Figure 6-2 demonstrates the left peripheral zone apical tumour which was missed by pathology. It shows high signal on DWI signal with correlating low signal on the ADC map and was recorded as a Likert 5 lesion (highly suspicious for tumour) by both radiologists.

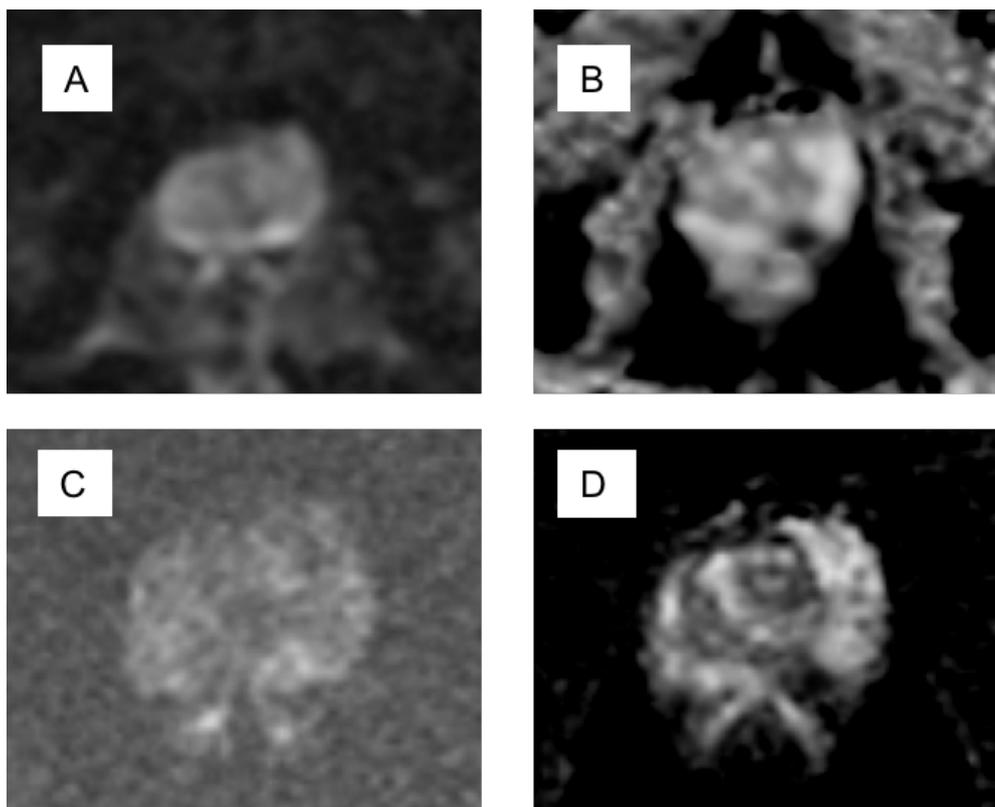


Figure 6-2. Example of a tumour missed on initial assessment by the histopathologists. A - cDWI b1500, B - cDWI ADC map, C - sDWI b1500, D - sDWI ADC map.

On another instance, a lesion was seen in the left hemi-gland by both radiologists, on all sequences, and the histopathology map showed bilateral tumours, but the right sided histopathology tumour was identical in shape and location to the radiologists' left sided tumour, but just mirrored to the other side of the gland. During the cutting of the prostate each quadrant of the outside of the gland is painted with a different colour so when the axial slices are reviewed later the pathologists know what anterior and posterior, and left and right are. On review of the TRUS biopsy in this case a Gleason Grade 4+3 lesion was found on the left and a 3+4 tumour found on the right, which was the opposite from the histopathology findings. In this instance, the pathology team believed there could have been error in the documentation of which colour paint corresponded to which side of the gland leading to sides being mixed. These are two examples of potential errors which were avoided, but there may be others which were not detected.

Further error could have been introduced whilst drawing ROIs of lesions which were not visible on DWI or T2WI images. This was done by KG, a radiology registrar with sufficient experience to report prostate MRI independently but less experienced than the readers. It was performed whilst assessing all MRI images, pictures of the histopathology slides, and the histopathologist reports. Due to the mismatch of number of pathology slides compared to MRI slices through the prostate, there is potential for incorrectly localising the tumour and ROI, particularly if the tumour is small and in a region of post-biopsy haemorrhage. This could have led to tumour ROI's containing some or no tumour tissue, leading to inaccurate results. Some studies limit the assessment to index lesions which increase the likelihood of correctly localising the tumour [98,203] and some studies only drew ROI's on tumours which are visible [82,157]. Given the multifocality of prostate cancer it was decided by the researchers to perform a quantitative and qualitative assessment of all tumours.

Correlating histopathology and radiology findings is technically challenging. Isebaert et al. [141] tried to formalise the registration of MRI and pathology slides. The diagram in Figure 6-3 demonstrates their methodology. This approach formalises the process the radiologists and histopathologists attempted to follow in this study. Other approaches to improve correlation and ROI placement include using an additional radiologist, separate from the readers and radiologist who draws the ROI, to decide the lesion location in the histopathology report and based on images of the slides [142,204]. Another possible way to improve the process is for the correlation for all lesions to be performed in consensus between the reporting histopathologist and a separate radiologist [81,168,178]. The use of a separate radiologist is to minimise bias as this radiologist should be blinded to the MRI reports of the reporting radiologist in the study.

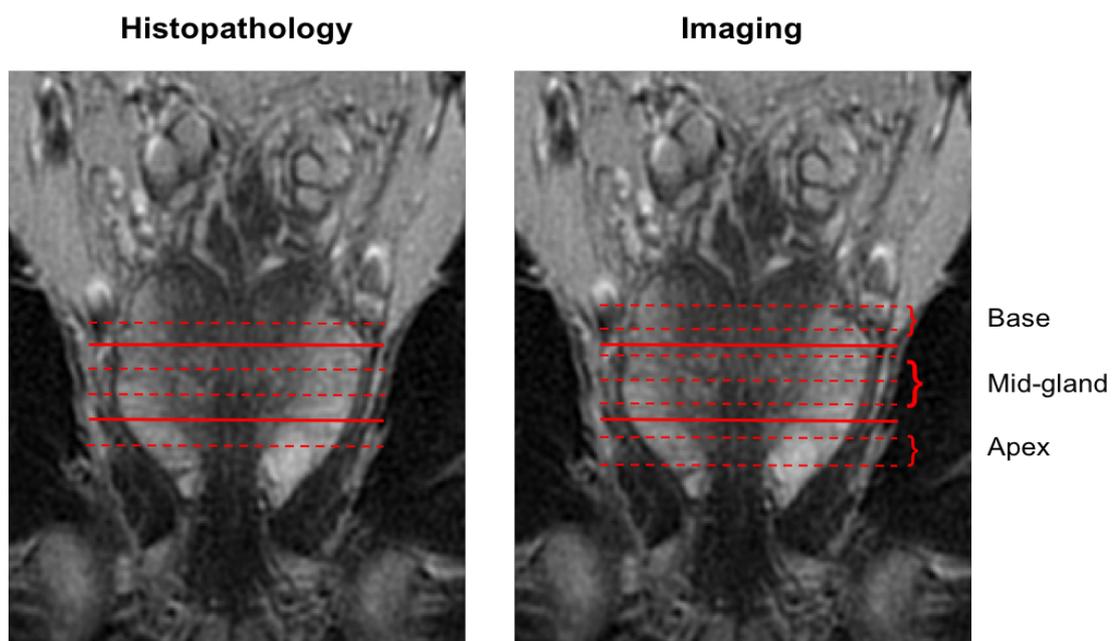


Figure 6-3. Diagram based on the methodology by Isebaert et al. (2012) [141] demonstrating the correlation of the different segments of the prostate on histopathology and radiology.

If this current study was performed again further steps could be introduced to improve the lesion-based accuracy assessment, and the correlation of prostate MRI and pathology. Firstly, the correlation could be performed in the presence of a histopathologist. Secondly, the lesion-based accuracy assessment could be validated by repeating the correlation assessment of lesions by the radiologist and histopathologist. Thirdly, prior to the initial reads of the prostate scans by the radiologists, a separate radiologist could have reviewed the scans and decided which slices correlated with the different height sectors of the gland. Finally, instead of hand cutting the prostate specimens, histopathologists can use a device which cuts the prostate at a consistent angle and at set intervals, which could be matched to the MRI slice thickness [127].

Reliability of DWI sequences and comparison to current literature

As well as being accurate a diagnostic test should be consistent and the precision of a test depends in part on levels of agreement between assessors. A weighted Cohen's kappa test was performed on the sectoral accuracy analysis of all 4 protocols. There were overlapping confidence intervals and high p-values between the comparable sequences indicating the differences were

not significant. However, the trend was for the cDWI sequence to be more reliable than sDWI regardless of the presence of T2WI. The most reliable test was with the cDWI sequence where the weighted kappa was 0.62 indicating substantial agreement [200]. When T2WI was added into the assessment of either DWI sequence the kappa value fell by 0.08 for cDWI and 0.07 for sDWI.

The sectoral based reliability was advantageous over the reliability based on lesion assessment as the numbers of sectors were greater than the number of lesions. Lesion reliability can be assessed by determining how many lesions are seen by both radiologists, and can be measured as percentage agreement. However, without true negatives the kappa coefficient cannot be calculated and as a result only sector-based reliability was performed. As the readers often started their assessment by identifying lesions and then drawing the lesion in the tumour map, a process they do in normal clinical practice, the information on the lesion-based reliability should be incorporated into the data and results of the sectoral approach. As with the correlation of pathology and radiology, a potential source of error in the sectoral reliability assessment is differences in interpretation between readers as to which slices represent the apex, mid-gland and base of the prostate. This is an important decision to make in clinical practice as prostate MRI is increasingly being used to guide biopsy and focal therapies rather than stage cancer.

Reader experience is important in the assessment of MRI prostate examinations [205]. In this study, we assessed the reliability and accuracy of two experienced readers. Both are uro-radiologists, have completed fellowships in the field, and much of their reporting work each week is reporting prostate MRI. It would have been useful to assess the effect of an inexperienced reader on the reliability. A third reader adds to the confidence the reliability results are generalisable, and the addition of an inexperienced reader could study whether prostate MRI is an examination more radiologists could report. Currently at NNUH amongst 32 radiologists, only 3 radiologists report prostate MRI regularly, in comparison to other MRI examinations such as MRI heads and lumbar spines which almost all the radiologists in the department report. The

specialised nature of prostate MRI examinations was the reason for performing the diagnostic accuracy study with two experienced readers, rather than an experienced and inexperienced reader.

There is a wide variation in the interobserver reliability between and within prostate DWI diagnostic accuracy studies. For example, Ohgiya et al. [85] found reliability to increase with increasing b-value with a kappa of 0.35 for T2WI and b500, a kappa of 0.55 for T2WI and b1000, and a kappa of 0.75 for b2000. However, Koo et al. [91] found reliability to increase from a kappa of 0.22 using a b300 ADC map, to 0.27 for a b2000 ADC map. The moderate and substantial level of agreement of sDWI and cDWI, respectively, is similar to many other diagnostic accuracy studies [83,141,180,184,204], but comparing them directly is not reliable given the different experience of readers, and technical parameters and combination of sequences used.

One difference between the published reliability of prostate DWI accuracy studies and the FODIP interobserver reliability is the improvement in reliability with the addition of anatomical sequences, typically T2WI alone compared to T2WI and DWI [83,178,180,184,204]. For example, Doo et al. [204] found reliability to rise from a kappa of 0.61 to 0.69 with the addition of ADC to T2WI. This study found reliability to fall with the addition of T2WI to DWI, a result which occurred with both DWI sequences. Only Morgan et al. [169] found a drop in reliability with the addition of DWI to T2WI (kappa of 0.51 with T2WI alone, and kappa of 0.33 with T2WI and DWI). A possible explanation for this is the manner in which T2WI was assessed. The T2WI sequence was used in protocols B and D and for each patient protocols B and D were read immediately after reading DWI alone in protocols A and C, respectively. This is in comparison to the time gap in reporting imposed between cDWI and sDWI reads to reduce recall bias. The decision to read the T2WI sequence for every case immediately after was intended to reduce the total time of the reads, but potentially introduced bias to the reports. The initial read of the DWI sequence could have had varying influence between readers when it came to reporting the T2WI and DWI sequences

immediately after. It would be beneficial to repeat the reads of the DWI, and T2WI and DWI sequences with a time period introduced in order to more reliably determine whether the reduction in interobserver reliability was indeed due to the immediate read with the T2WI sequence after the DWI before moving on to the next case.

Small-field of view results compared to current literature

Currently the only study which has assessed the qualitative diagnostic accuracy of a small field of view DWI technique at detecting prostate cancer is by Brendle et al. [119]. In addition, to finding an improvement in lesion conspicuity and reduced distortion they found a sensitivity of 0.66, specificity of 0.99 and an AUC of 0.82. These diagnostic findings were not significantly different from the conventional imaging they used. The readers, who were blinded to clinical data, read the MRI in consensus and had T2WI and T1WI sequences available. The study was of 15 patients using a 12 sector reporting template and had prostatectomy specimens as a reference standard. It is unclear if they assessed just for index lesions or allowed for multifocality and whether they used a Likert scale or had a binary cut-off. The sensitivity is higher than found in the FODIP study (0.44–0.52 depending on Likert cut-off and reader), as is the specificity (0.88–0.95) and AUC (0.79 and 0.71). The reason for the difference is perhaps due to the imbalance of sectors with tumour to those without (29 with vs 151 without) in Brendle's study suggesting they assessed only for the index lesion. Furthermore, the assessment of the studies in consensus by two experienced readers could account for the differences.

The group also used b values of 50 and 800 sec/mm² for their ADC maps which was different from this study. This factor's influence is also supported by the higher mean and minimum tumour ADC values when compared to the FODIP study (ADC_{mean} of 1000 x10⁻⁶ vs. ADC_{mean} of 766 x10⁻⁶ mm²/sec; ADC_{min} of 700 x10⁻⁶ mm²/sec vs. ADC_{min} of 500 x10⁻⁶ mm²/sec). This difference is striking, particularly if they did only assess index lesions which tend to be of higher grade and lower ADC value. Feng et al. [115] used the same small FOV DWI sequence as the FODIP study (FOCUS), but

used b-values of 0 and 800 sec/mm², had a pre-biopsy scan, used TRUS biopsy as a reference standard, and qualitatively assessed image quality parameters and quantitatively assessed the ADC values of index lesions. Like Brendle et al. [119] they found qualitative improvements in image quality using the small FOV DWI sequence, and the ADC value of tumours was higher with the small FOV technique in comparison to the large FOV sequence with comparable qualitative image quality results between studies.

It is believed that by reducing the field of view there is a reduction in partial voluming leading to a more accurate assessment of the ADC value. The large discrepancy in tumour ADC between the FODIP study and others in the literature is potentially due to the technical parameters chosen. It is also likely that the difference in tumour ADC between small FOV and conventional DWI sequences in the FODIP study and the other small FOV DWI studies is related to other factors beyond FOV, such as choice and number of b-values used to create the ADC map and the SNR. The reduction in the FOV leads to lower SNR [206]. The use of small FOV techniques at 3T is to counteract many of the susceptibility effects at the higher field strength, but when combined with higher b-values, which also lowers the signal, the benefit of greater signal at 3T may not be enough to prevent the loss of signal from reducing the FOV and increasing the b-value. The result of a reduction in SNR can be an underestimation of tumour ADC [207,208], which could account for the lower ADC values of tumours in this study.

Accuracy results of DWI sequences compared to the meta-analyses and current literature

The results of the sectoral diagnostic performance of DWI alone compared to the qualitative accuracy analysis in chapter 2 (4 studies [90,91,141,142]) has shown this study to have a reduced AUC (0.76 for cDWI and 0.72 for sDWI vs 0.95 for high b-value meta-analysis), a similar sensitivity compared to the pooled sensitivity (0.62 for cDWI and 0.45 for sDWI vs 0.58), and a worse

specificity (0.85 for cDWI and 0.92 for sDWI vs 0.93). The reduced sensitivity of the small FOV DWI sequence compared to the pooled results is large and suggests that as a sequence it missed a significant number of tumours which may outweigh the benefit of its high specificity. The difference in AUC between this study and the meta-analysis is large, but needs to be interpreted with caution given this study AUC is derived from a ROC curve created from the Likert scores and the meta-analysis from a summary ROC curve, which is a different process that is prone to errors, for example, from sampling variability between studies and bias in the weightings in included studies.

Comparing the results of this study to the meta-analysis is challenging given the heterogeneity between the included studies. For example, even between the four studies there was a difference in the number of sectors used (between 10 and 24), field strength, b-value, use of binary or Likert score, and which combination of DWI and ADC images were used. The closest comparison study is that of Koo et al. [91] which was matched for field strength and visual assessment of DWI and ADC together, but used 10 sectors and was comparing b1000 and b2000 images. Like the pooled results in the meta-analysis their sensitivity was much higher (0.85 at b1000 and 0.74 at b2000) and their specificity comparable (0.94 at b1000 and 0.96 at b2000). One difference between the FODIP study and the four in the visual assessment subgroup is the higher percentage of sectors containing tumour (almost half) in the FODIP study. For example, in the studies by Koo et al. [91] and Kim et al. [90] only approximately 25% of sectors contained tumour. However, Isebaert et al. [141] described 55% of sectors being positive and their results were similar to this study (Sensitivity 0.37, Specificity 0.94). The eligibility criteria for the presence of tumours in these four studies appears similar to the FODIP study and they all tested for multifocal tumours. Potentially, if the difference between the number of tumour positive sectors is due to an omission of smaller lesions this could account for the worse sensitivity experienced in this study.

There is greater heterogeneity between the studies of the T2WI and DWI meta-analysis (chapter 3, page 82). There was no included study which matched the FODIP study for high b-value ($b \geq 1000$), use of DWI and ADC map, 12 sector assessment of the whole gland, 3T field strength, and with radical prostatectomy specimens as reference standard. There were many which contained most of the same study characteristics.

For reader 1 with a Likert cut-off of 3 and above considered positive for tumour; the sensitivity, specificity and AUC results for conventional DWI alone were similar to the results of the results of DWI and T2WI combined meta-analysis, but there was a significant improvement of reader 1's sensitivity and AUC when T2WI was added to the small FOV DWI sequence assessment. In general, the pooled sensitivity of the T2WI and DWI meta-analysis are more like the results of the FODIP study than to the high b-value DWI meta-analysis. The T2WI and DWI meta-analysis pooled sensitivity was 0.68 (reader 1 cDWI and T2WI - 0.60, sDWI and T2WI - 0.50), pooled specificity was 0.84 (reader 1 cDWI and T2WI - 0.85, sDWI and T2WI - 0.93) and AUC was 0.84 (reader 1 cDWI and T2WI - 0.79, and sDWI and T2WI - 0.79). When compared to matching subgroups such as 3T, radical prostatectomy and visual assessment of both DWI and ADC maps, the difference between the lower sensitivity found in FODIP study and those subgroups increases, but specificity remains similar.

Most of the sectoral-based accuracy studies in the T2WI and DWI meta-analysis had a smaller proportion of tumour positive sectors compared to this study. There were a few which had greater than 40% of tumour involved sectors [83,168,169,177,180,209]. Ueno's two studies of high b-value DWI and computed DWI images alone against a radical prostatectomy specimens using 8 sectors per patient at 3T [83,168] found sensitivities ranging from 0.76 to 0.84 depending on b-values ($b1000$ and $b2000$), but their specificities were substantially lower (0.62 to 0.70). This perhaps was due to the different number of sectors used in their study, as by using 8 sectors and dividing the entire transitional zone into 2 sectors, rather than the 6 in the FODIP study, more

weight is given to the peripheral zone assessment in the results. Transitional zone tumours are harder to identify than peripheral zone tumours with increased false positives because of the overlap in appearance with hyperplastic nodules, typically located in the transitional zone, with tumour [81].

Another limitation of comparing the FODIP study's results with the available literature and a possible explanation in the discrepancy between results is publication bias. It is an important form of bias when studies with favourable results are more likely to be published than those without [210]. In addition, they are published earlier and within journals with higher impact factors. Studies with significant differences are more likely to be published than ones without significant findings. Often diagnostic test accuracy studies do not require ethical review leading them to be harder to trace results and follow up [211]. Funnel plots have been demonstrated to be misleading for diagnostic studies with low power [191] and therefore, was not performed in the meta-analyses. Nevertheless, it should be considered as a potential cause for the difference between results.

The number of published studies which have assessed the diagnostic performance of DWI at detecting prostate tumours by following a lesion-based approach is much smaller than the sectoral-based approach. The use of the radiologist reporting templates and prostate map requires the reader to draw a lesion in a specific area of the prostate and also detail other characteristics of the individual lesions in the adjacent table. This facilitated the simultaneous assessment of lesion and sectoral accuracy. The reporting instructions for the radiologists (Appendix 2) was designed to mimic normal clinical practice as much as possible as a reporting radiologists will probably assess the MRI firstly for the presence of lesions, then draw the lesion in the prostate map, and finally consider the remaining sectors to determine the likelihood of tumour absence. This last step is furthest from a radiologists normal practice, and therefore, the lesion-based approach could be considered the closest to daily clinical practice of radiologists. But

the sectoral-based approach is popular in the published literature as it provides a complete 2x2 contingency table and facilitates the creation of ROC curves.

There is variability among the published lesion-based accuracy study methodology with some studies assessing for index lesions [81,135,136] and some for all lesions [99,212,213]. There is also variability among the choice and combination of sequences assessed, choice of b-values, and blinding of radiologists to histopathology results. Localising tumours to specific areas of the prostate is important as the recent introduction of focal treatments, such as brachytherapy, high-intensity focused ultrasound and cryotherapy, has necessitated the need for diagnostic imaging to extend beyond staging tumours and provide more detailed information regarding size and location of lesions. In comparison to Rosenkrantz's [136] assessment of lesion-based accuracy of high b-value DWI sequences, including both acquired and computed images, the b1500 DWI subset had a lesion detection rate of 77% for index lesions with Gleason Score $\geq 3+4$. In comparison to the results of the both DWI protocols in the FODIP study, reader 1 identified 82.5% of index lesions with the conventional DWI sequence and 75% with the small FOV DWI sequence. This perhaps highlights the advantage of having the ADC map as well as the b-value images when assessing for index lesions.

In the FODIP study, although, the inclusion of false positive results in the accuracy assessment for index lesions due to the readers' requirement for assessment for multiple lesions at the outset the evaluation for index lesions has merit. The consensus in the current literature places increased importance on the index lesion, rather than all tumour foci, and urologists are now using the information about the index lesion to provide prognostic information and drive management decisions [36,214,215]. The high detection rate of index lesions by conventional and small FOV DWI is promising and supports the use of higher b-values. These results in combination, with the high PPV for both these sequences, in the assessment of all lesions, adds weight to the

potential for treating patients' prostate cancer on the basis of a suspicious finding without biopsy, as a high PPV is necessary to avoid focally treating benign tissue whilst missing the index lesion.

In summary, the FODIP study sectoral analysis results of the small FOV and conventional DWI sequences are difficult to compare directly to the published literature and the performed meta-analyses given the lack of studies with matching study protocols. In general, the results demonstrate that the sensitivity results of both sequences are below that of the published literature, and substantially for the small FOV DWI sequence. However, the specificity results are comparable, with the small FOV DWI sequence perhaps slightly better. The difference in the number of positive tumour sectors between the FODIP study and the majority of other comparable diagnostic accuracy studies is a potential cause for the discrepancy.

Comparison of b1000 and b1500 results

When quantitative tumour characteristics were compared at b1000 and b1500 there was a significant increase in SNR of b1000 over b1500 for both DWI sequences. Conversely, the tumour contrast was significantly improved with b1500 for both DWI sequences. As the b-value increases the rate of reduction of signal within non-cancerous tissue decreases more quickly than the signal loss within cancerous tissue; this leads to the improved contrast. However, when the b-value becomes too high the benefit of the added contrast between cancer and non-cancerous tumour is lost due to the lack of signal within the image. Multiple studies have shown SNR to be significantly lower at b2000 compared to b1000 [87,88,195], but only a few studies have compared SNR of b1500 and b1000. Similar to the FODIP study, Metens et al. [87] found higher SNR with b1000, however, Wang et al. [88] found the opposite. The SNR values found in those two studies were similar to the values found in small FOV DWI sequence (between 16 and 20), but the SNR of the conventional DWI sequence was 25.7 and 23.16 for the b1000 and b1500 sequence, respectively. The cause of the high SNR of the conventional DWI sequence is unclear as most of the imaging parameters that influence signal was comparable with Wang et al., such as FOV and TE. It is

possible that the number of excitations were higher for the conventional DWI sequence, but this was not published by this group. The most likely explanation for the higher SNR in the conventional DWI sequence is the larger FOV which results in a decrease in signal noise. The improved contrast parameters at b1500 over b1000 for both sequences is consistent with other studies comparing these values [87,216] and justifies the choice of using b1500 as the highest b-value when optimising both DWI sequences prior to the commencement of recruitment.

ADC value of tumours

As in most studies, the ADC value of tumour was significantly lower than normal prostate tissue. Using 3 b-values the ADC values of tumour using a small FOV sequence were significantly lower than the large FOV conventional sequence. Despite the reduction in ADC value of tumour on sDWI there was a significantly greater AUC with cDWI when a ROC curve of mean ADC values was performed. The AUC for cDWI and sDWI were 0.96 and 0.90, respectively, which indicates ADC value measurement has excellent accuracy at distinguishing tumour from non-tumour tissue. With cDWI an ADC cut-off of 1153 mm²/sec gave a sensitivity and specificity of 0.86 and 0.98, respectively, whereas with sDWI an ADC cut-off of 898 mm²/sec gave a sensitivity of 0.77 and specificity of 0.93. However, there should be some caution when interpreting these results given the possible mis-registration of the ROIs drawn on lesions to measure tumour ADC values, particularly as lesions were less visible and sensitivity lower for the small FOV DWI sequence.

The results of the quantitative ADC AUC results are substantially higher than the comparable subgroup in the high b-value meta-analysis (Chapter 2, page 74). This was potentially because ROIs were drawn with access to the histopathology reports (non-blinded) and with a optimised ADC value determined, as opposed to blindly with a pre-determined ADC cut-off. Furthermore, the ROIs were drawn around the margins of the tumour as seen on imaging rather than on the histopathology maps which were often larger and could have falsely lowered the ADC of tumour.

The ability to distinguish tumour from non-tumour using ADC values is useful, particularly given the higher AUC from the quantitative ROC curve compared to the qualitative ROC curve. But, the ADC values of tumours vary significantly in the literature. The choice and number of b-values have an influence on ADC value of tumours with ADC values falling with rising b-values [89,91,107]. However, this finding is not reliable as Kim et al. [90] found tumour to have a significantly higher mean ADC value with a b2000 ADC map compared to a b1000 ADC map (1580×10^{-6} vs 1190×10^{-6} mm²/sec, respectively). Even when studies with relatively matched DWI technical parameters the ADC values of tumour can vary such as between Kumar et al. [149], who used 5 b-values and a maximum b-value of 1000 sec/mm², and deSouza et al. [217], who used 4 b-values and a maximum of b-value of 800 sec/mm², the former study found tumour ADC to be much lower than the latter (980×10^{-6} vs 1300×10^{-6} mm²/sec). The variation in ADC values of tumour and non-tumour tissue between different DWI sequences and studies is the major reason why the quantitative measurement of lesions and regions of the prostate is not used in routine clinical practice. However, in individual institutions it may be useful to use 'ideal' ADC cut-offs between tumour and non-tumour tissue. These results indicate the possible reduction in ADC value of tumour when using smaller FOV DWI sequences with high b-values.

Correlation of Gleason Score and tumour ADC value

The mean and minimum ADC value of tumours had an inverse correlation with rising Gleason Score, which is in keeping with most other studies [104–106,218–220]. Only the minimum ADC value of the conventional DWI sequence had an almost significant inverse correlation with increasing Gleason Grade Group, but the Spearman Rank correlation coefficient was only -0.210, which is considered a weak correlation. The mean ADC value of tumours on cDWI and sDWI and the minimum ADC value on sDWI were not significant and the correlation very weak. Due to the small numbers of higher grade tumours, when splitting tumours into a Gleason Score 3+4 group and those more aggressive, the correlations were no more strong or significant.

This is in comparison to most other studies in which the correlation ranges from weak ($\rho = -0.32$) to fair ($\rho = -0.50$). The reason for the weaker correlation is not clear. Firstly, there were no Gleason 3+3 tumours included in the study which is unusual given that other studies, some with similar size eligibility criteria of 0.5cm in maximum axial diameter, have at least one Gleason Score 3+3 tumour and in general range from 2-35% of tumours in radical prostatectomy analyses [81,88,98,140,141,152,164,180,221]. The theory of decreasing ADC value and increasing restricted diffusion of tumours with increasing grade is based on the architecture of higher grade prostate tumours, which have a higher level of cellularity and therefore more restriction of water movement and lower ADC values. Some Gleason Score 6 lesions may have facilitated a stronger correlation as having Gleason Score 6 tumours with low levels of cellularity could have provided further data-points for which to add strength to the correlation. This, however, does not negate the evidence of the overlap of mean and minimum ADC values of tumours with different Gleason Scores in the FODIP study.

Another potential reason for the weak correlation found in this study compared to the wider literature is the method and reproducibility of drawing of the ROI. Care was made to avoid adjacent non-tumour tissue, but if the ROIs had overlapped this could lead to increased and more variable ADC values with more overlap and weaker correlation with the degree of cellularity of the tumour. By repeating the ROI drawings and validating this method of sampling with intraclass correlation coefficients a better understanding of the repeatability and accuracy of the ROI measurements would be possible.

Furthermore, many studies of ADC correlation with Gleason Grade were non-blinded and use index lesions. The index lesion of the tumour tends to be the largest, the highest grade and the most likely to have spread beyond the prostate. There are many studies which demonstrate a stronger correlation between ADC and Gleason score at radical prostatectomy than this study [104,106,157,183,218–220], but the following used just the index tumour [106,157,220], which

naturally selects the tumours which are largest, easiest to delineate and potential to get the most accurate ADC value from. Whilst, in this study ROIs were drawn on all tumours which met the eligibility criteria, some were not visible on imaging even with photographs of the tumour slides and the histopathology reporting template for guidance. In addition, the pathology slides were cut with a free hand and not necessarily on the same axial plane as the MRI therefore adding to the risk of misplacement of the tumour ROI.

The lack of Gleason 3+3 tumours is unfortunate as it is important clinically to be able to discriminate between low grade disease and intermediate and higher grades and to be able to do this non-invasively with imaging would be beneficial. Another important clinical distinction is between Gleason Score 3+4=7 and Gleason 4+3=7 as patients in the lower grade group often are offered active surveillance as opposed to more radical treatment for the higher-grade group. For this reason, ADC value was correlated with Gleason Grade Group rather than total Gleason Score which has been done frequently before. Gleason grade group 2 tumours equate to Gleason Score 3+4 and Gleason Score 4+3 are grade group 3. There are currently no MRI diagnostic accuracy studies in the literature which describe prostate tumours in terms of Gleason Grade Group, nor do any correlate grade group with ADC. As in all tumours and grade groups there was no significant difference in mean tumour ADC and the Gleason grade and nor was there between Gleason grade 3+4 tumours and those more aggressive. Part of the reason for sampling all tumours more aggressive than 3+4 was to increase numbers and to determine whether there was a significant difference between the two groups which is useful clinically. However, the significant overlap of ADC values and low level of correlation is in keeping with Rosenkrantz et al. (2015) [218]. In comparison, Vargas et al. [104], Nagarajan et al. [106] and Hambroek et al. [80] all demonstrated a significant difference in ADC values between these groups.

In addition to validating the method of drawing the tumour ROIs, investigating why there was a lack of Gleason Score 6 tumours, and assessing just the index lesions, there are other ADC value

related factors which can be employed to correlate with Gleason score. The variation in ADC values due to technical factor such as number and choice of b-values used to make the ADC map, can be minimised using ADC ratio. This involves dividing the ADC value of tumour by a constant within the field of imaging. Some groups have attempted to normalise the ADC ratio by dividing the ADC value of tumour with the ADC of urine within the bladder [177] and muscle [222], which can be highly variable and low signal respectively. Barrett et al. [203] used an ADC ratio of tumour to normal prostate tissue and demonstrated a stronger correlation with tumour grade than absolute tumour ADC and also a more consistent correlation with tumour grade using ADC maps created with different b-values.

Influence of post-biopsy haemorrhage on results

Post-biopsy haemorrhage, as denoted by high signal on T1 weighted images, was present in many of the prostate MRI studies and was seen in over one quarter of the lesions included in the analysis. 81% of prostates have haemorrhage up to 3 weeks after biopsy, just under half of prostates have haemorrhage after 3 weeks, and the haemorrhage can be detected up to 4 months after biopsy [70]. Haemorrhage within the tumour itself occurs in approximately 5% of cases [73], but the number of lesions containing haemorrhage at histopathology was not assessed in the FODIP study. Haemorrhage is believed to resolve more quickly in tumour tissue than in normal prostate tissue which gives rise to the Haemorrhage Exclusion Sign. This imaging sign occurs when low signal tumour is outlined by high signal haemorrhage on T1WI [69] and can often aid the radiologist in detecting tumour in the presence of haemorrhage. It occurs in 20% of patients with post-biopsy haemorrhage, and has a positive predictive value of just over 50%. In this study, the radiologists were not asked to comment on its presence, but their access to T1WI on all the imaging protocols may have resulted in the detection of tumour and influenced the accuracy of the imaging results.

The results of the qualitative assessment of accuracy indicated that both radiologists' ability to detect tumour improved when assessing for tumours with no peri-lesional haemorrhage compared to those with. This finding was consistent regardless of radiologist, diffusion sequence, or presence of T2WI sequence. An example of tumour seen outwith haemorrhage and missed within an area of haemorrhage is demonstrated in the figure below.

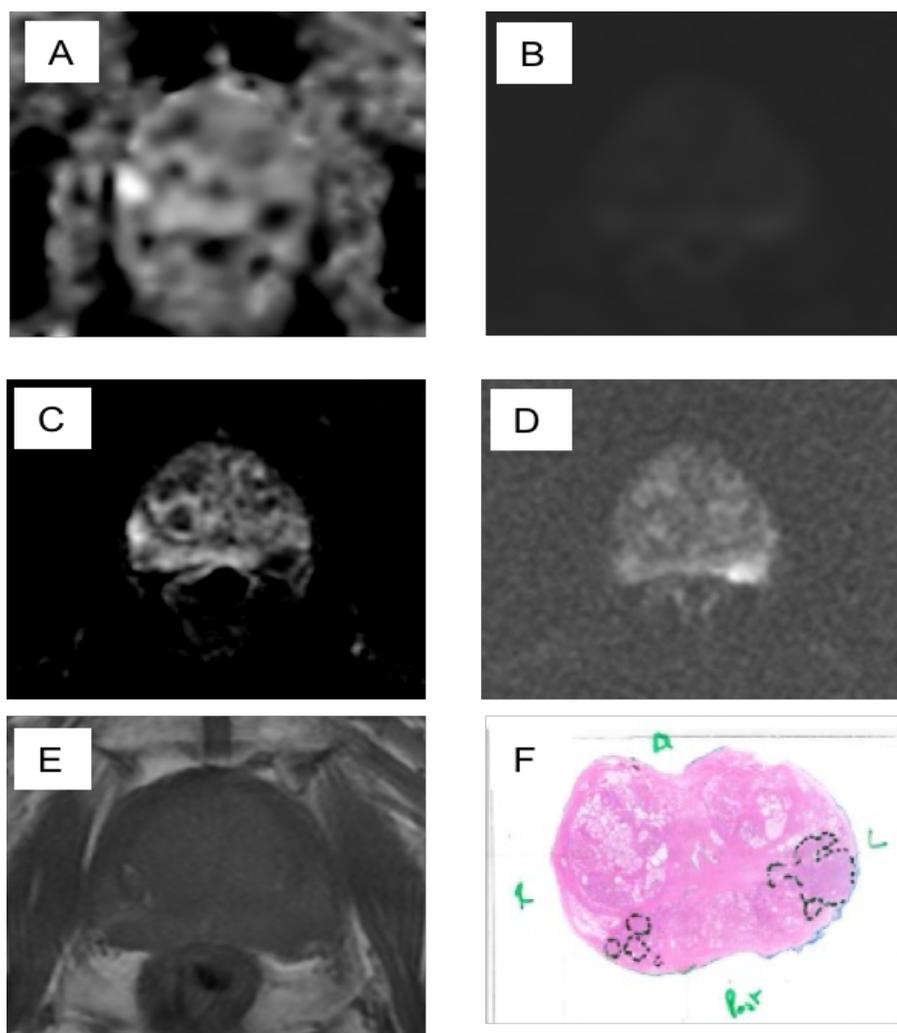


Figure 6-4. Example of peripheral zone tumours. The left PZ tumour was detected. The right PZ tumour in an area of haemorrhage was not seen. A) cDWI ADC map, B) cDWI b1500, C) sDWI ADC map, D) cDWI b1500, E) T1WI , F) Correlate pathology slide.

The quantitative assessment of haemorrhage in tumours resulted in a significantly higher mean and minimum tumour ADC with the conventional DWI sequence and mean ADC with the small FOV DWI sequence. A rise in the ADC value of tumour tissue compared to the adjacent prostate tissue, which is of higher signal, can reduce the tumour to normal tissue contrast potentially

reducing the ability of the reader to detect the tumour. The tumour CR and CNR was significantly greater in lesions with no haemorrhage with the sDWI ADC map which would suggest the rise in the ADC of haemorrhagic tissue is possibly a cause for the decreased lesion detection. However, the alteration in contrast was not seen in the conventional DWI ADC map, nor in either DWI sequence b1000 or b1500 imaging set, suggesting other causes for decreased accuracy. The AUC for non-haemorrhagic lesions using ADC was greater in lesions without haemorrhage compared to all lesions with sDWI and cDWI again indicating a potential benefit of not having haemorrhage in the prostate when assessing with MRI.

The current practice at NNUH is to wait 4 weeks following biopsy for MRI. This delay accounts for a considerable amount of time in the National referral to decision-to-treat cancer target. For a period early in the recruitment window (September 2015-November 2015) the 4 week delay ceased with some patients being scanned within 1 week of biopsy. Although, this goes against guidelines for delaying imaging after biopsy, which suggest a delay between 3 and 10 weeks [68,223], the effects of post-biopsy haemorrhage using multi-parametric sequences, such as DWI, is believed to be much less than T2WI and not significant [72,73,224]. Rosenkrantz et al. (2010) demonstrated that there was still a significant difference in ADC of tumour tissue compared to haemorrhagic and non-haemorrhagic peripheral zone tissue [73]. The qualitative assessment of sectoral accuracy by the same group demonstrated that none of the diagnostic parameters were affected by haemorrhage using DWI and DCE, but T2WI demonstrated significantly decreased sensitivity when haemorrhage was present in the sector [72].

A cut-off delay of 3 weeks from biopsy to MRI was decided for the FODIP study. A decision to lower the delay from the local 4 week guideline was made as the study's primary objective was to assess DWI sequences, which are less affected by haemorrhage. In addition, Park et al. [225] demonstrated accuracy of mpMRI was not significantly different in MRI performed before and after 21 days. Our results, however, do suggest post-biopsy haemorrhage can influence accuracy

and therefore, justifies a cut-off delay being implemented. There is an inverse correlation with the visual measurement of post-biopsy haemorrhage and time following biopsy [72,225] and 21 days is cited by multiple studies as a key time where the haemorrhagic effects are reduced sufficiently to reliably use MRI and therefore this was the justification for choosing this delay [70,226].

The results demonstrate an effect of haemorrhage on qualitative and quantitative accuracy of the ADC maps of both DWI sequences. These results should be treated cautiously as the lesions which were deemed to have peri-lesional haemorrhage may also be more likely to be of lower Gleason Grade or smaller, both of which affect accuracy. A multivariate logistic regression analysis might have been a useful way to determine the combined effects of each of these factors on accuracy and could have provided information on the importance of each.

Current trends in prostate cancer investigation is moving towards imaging patients before biopsy. Following the PROMIS trial results indicating the strength of mpMRI and weakness of TRUS biopsy at detecting significant cancer when assessing the presence of tumour on a whole gland basis (unlike sectoral and lesion based accuracy assessment in the FODIP study) [58], the shift to a pre-biopsy MRI is being implemented in more centres in the UK. This has changed the focus to detecting clinically significant cancer rather than all cancers, guiding biopsy rather than staging disease, and with the main benefit being reduced number of TRUS biopsies. This approach also has the benefit of reducing the effects of post-biopsy haemorrhage on assessment of tumour and is an argument for assessing the potential diagnostic accuracy of lesions and sectors which are free from the effects of haemorrhage.

Further limitations

To reduce recall bias there was a 2 week gap between reads of the cDWI and sDWI sequences. However, the T1WI and T2WI sequences were present in both the reads, which may increase the chances of recalling the case. The order the cases were assessed in was randomised using an

online number randomiser [194] to reduce the impact of recall bias. Although, 40 cases in total were assessed this was done in two blocks of 20. Initially cases 1-20 of cDWI were assessed, followed by cases 1-20 of sDWI, then 21-40 of cDWI, and finally 21-40 of sDWI. This was due to a slowing of the recruitment process in September 2016 where it became apparent that cases 36-40 would take another 4 weeks to acquire as the radical prostatectomy procedures were not scheduled until then. Therefore, in consensus, the research team decided to commence the blinded radiology assessment on the first 20 cases and by halving the groups this minimised the impact of the increased risk of recall.

The study was a prospectively planned and ethically approved observational study. The observational nature of the FODIP study did subject it to potential selection and verification bias. There was careful planning of the study before the recruitment window opened. Meticulous consideration of the technical parameters of the DWI sequence was made in discussion with an application specialist from General Electric, the manufacturer of the 3T scanner. These sequences were tested and implemented before recruitment opened in September 2015. In addition, the inclusion and exclusion criteria was set prior to commencement, and a detailed plan was made to identify potential patients eligible for the study. To gain ethical approval without the need for informed written consent from patients it was important to emphasise that at no point was this study going to impact in any patient's clinical management pathway and therefore a significant amount of prospective planning occurred. This greatly reduced the need to alter any aspects of the methodology other than the above mentioned splitting of the case load in half and the insertion of a minimum of 21 days post-biopsy to the eligibility criteria, the limitations for which have been discussed. However, the patients were not prospectively recruited as the study proceeded, with identification of potential eligible subjects after prostatectomy had been performed. Therefore, during the routine NHS clinical care of the included patients the results of the assessment of the patient's MRI, and also the histopathologists interpretation of biopsy specimens, and surgeons and anaesthetists assessment of appropriateness for surgery, all

influenced whether the patient proceeded to radical prostatectomy and therefore introduced verification bias into the study.

Finally, the use of radical prostatectomy specimens as the reference standard inserts a significant selection bias to the study. This choice of reference standard is thought to be gold standard as it allows a more accurate and reliable assessment of the entire gland and correlation with axial slices of MRI sequences. However, a specific group of patients are selected for the procedure. They tend to have tumours of higher grade (Gleason Grade Group 2 and above), but of a locally confined stage such as T2a, T2b, T2c and early T3a disease, with no significant nodal spread. In addition, they are younger and fitter patients as the surgery does have a significant morbidity. When it comes to the interpretation of the MRI in this group, prostate cancer is present, as urologists would not proceed with the surgery without histological proof, and the lesions are more likely to be visible as they are of higher grade. The diagnostic accuracy of MRI in a broader clinical setting is possibly going to be worse as many more low-grade and non-visible tumours will be present in this population group. However, in routine clinical work the radiologist typically has access to PSA, digital rectal examination, and if in the post-biopsy population, the results of this investigation, compared with the readers in this study who were blinded to this information.

The other potential reference standard which allows sampling of the whole gland is template prostate mapping biopsies as in the PROMIS trial [58], which carried out template biopsies of the entire gland at 5mm intervals. However, PROMIS assessed the likelihood of tumour presence on a whole-gland basis, but did not aim to localise tumours within the gland nor did they assess for the tumour stage. The use of this technique would allow the assessment of MRI earlier in the diagnostic pathway, including some men who do not have cancer, however, its use is limited to large trials, with significant financial support, large multi-disciplinary infrastructure, and written consent from patients. Furthermore, its ability to correlate with MRI in the localisation of tumour within the gland has not been tested.

There are advantages to using a technique and a methodology which has been used by most the published literature. The repeated use of radical prostatectomy specimens and the relative ease of correlation with MRI has been tested, optimised and to some extent validated. In addition, by following set methodology it allows a more confident comparison of this study's results with the published literature and the meta-analyses performed.

Wider implication of results and potential future research

This is the first study which has assessed the diagnostic performance of small FOV DWI using radical prostatectomy specimens as the reference standard. Additionally, it's comparison with the conventional large FOV has not been performed in this population. Finally, no study has combined the small FOV technique with higher b-values.

The strength of the study is the combined sectoral and lesion-based accuracy assessment with a statistical comparison of results, the reliability analysis, and the quantitative tumour characteristics. Its weaknesses are the known difficulties in correlating MRI with radical prostatectomy specimens, which could have been improved by the process being performed by a 3rd experienced radiologist in consensus with a histopathologist. Additionally, the potential influence of haemorrhage on the interpretation of the results is unknown, but could have been significant.

From the results, the small FOV DWI sequence is significantly poorer in the detection of prostate cancer than the conventional large FOV sequence, but conversely, its strength is in its ability to rule out prostate cancer by minimising the overcalling of lesions. The small FOV DWI sequence is moderately reliable, but the cDWI is substantially reliable. The results of both sequences are worse than the published literature and the meta-analyses performed in Chapters 2 and 3. The reasons for this are unknown, but may have been influenced by post-biopsy haemorrhage, or radiologist and histopathologist inexperience with the research methodology. Publication bias is

another potential cause for this discrepancy. The ADC values of tumours on the small FOV DWI sequence were significantly lower than the conventional DWI sequence, which is out of keeping with other studies using small FOV at lower b-values suggesting the combination of small FOV and high b-value are the cause of this, potentially due to the loss of signal combining these techniques. Finally, the correlation of tumour ADC value with Gleason grade was weaker with the small FOV sequence, but both sequences demonstrated a weaker correlation compared to other studies.

With these results it is unlikely, with the technical parameters used, that the sDWI sequence would replace the cDWI sequence locally given its poor lesion detection rate. It is debateable whether it would be used alongside the conventional DWI sequence as it takes approximately eight minutes to complete the sequence which is considerable in a busy radiology department. Shortening the imaging protocol has the potential to increase capacity. As well as the improved diagnostic performance, the larger FOV also provides diagnostic information about the entire pelvis which allows the reader to assess for bony metastases, localise lymph nodes, and potentially identify other pathology not seen on the small FOV. The small FOV DWI sequence used is commercially available, but software and sequence upgrades can be considerable in cost, and if the potential benefits are minimal then it may not be worth considering acquiring. However, the small FOV DWI sequence can be used on other body parts, on which this sequence should be validated.

There is potential for further work to be done on the images obtained and data collected as part of the FODIP study. A logistic regression analysis of the accuracy data with haemorrhage, tumour size and gleason score as independent variables would be a useful analysis. It is important to better understand the reasons for potentially spotting and missing lesions, as this could lead to retraining of radiologists to assess for variables which have a large effect on accuracy thus leading to improved diagnostic performance.

Although, the diagnostic performance of sDWI was poorer than cDWI, anecdotally amongst the prostate radiologists at the NNUH there is a preference for the small FOV imaging. There are potential reasons for this which could lead to further research. Performing image quality analysis: qualitatively with subjective assessment of distortion and artefact, and quantitatively doing tumour texture analysis and using software to measure distortion could allow a better understanding of small FOV DWI's benefits and limitations.

Further research should be performed on the sequence as the medical physics theories underpinning the use of small FOV DWI sequences on the prostate are sound. A potential limitation has perhaps been the combination with high b-values due to the loss of signal. Therefore, it would be useful to assess the accuracy and reliability of the sequence at lower b-values. Additionally, ADC maps can be created using different combinations of b-values allowing qualitative diagnostic performance, quantitative analysis, and correlation of ADC value and tumour grade to be performed resulting in better optimisation of the sequences used.

The creation of ADC maps with b-values greater than 1000 sec/mm^2 is controversial and a potential way to gain the diagnostic benefits of higher b-value, but without the loss of signal is to use a computerised b-value. This allows the creation of a synthetic high b-value image by extrapolating the data acquired in the lower b-values, for example 100 and 1000 sec/mm^2 . This provides a high b-value image without the associated loss of signal and provides an ADC map derived from recommended b-values. Finally, extending the study to include more patients would increase the confidence in the findings and potentially reveal more significant differences between large and small FOV DWI sequences.

Chapter 7 Summary of FODIP study

The diagnosis and management of prostate cancer is a challenge. The spectrum of severity of disease, the non-specific nature of the PSA blood test, and the high false negative rate of the most common biopsy technique contribute to this. MRI is continually developing with improvements in diagnostic performance. In particular, there have been notable recent advances in DWI imaging with the use of high b-values, which is now established practice, and the reduction in the field-of-view.

A diagnostic test accuracy meta-analysis of high b-value DWI alone for detecting prostate cancer included 10 studies and on which the quality of studies was considered good. The results demonstrated that a visual assessment of high b-value DWI images was significantly more accurate than quantitative accuracy using ADC measurement for tumour.

The major limitation of the high b-value meta-analysis was the small number of included studies, particularly those in which the primary aim was to assess high b-value DWI. Most diagnostic accuracy studies of DWI use T2WI as well, with the latter sequence seen as a standard presence and the technical parameters rarely discussed. The meta-analysis of diagnostic test accuracy of visual assessment of T2WI and DWI for the detection of prostate cancer included 22 studies and found this technique to be a good diagnostic tool and the best diagnostic performance was with b-values greater than 1000 sec/mm^2 and assessment of the ADC maps.

Most DWI techniques scan the entire pelvis with the prostate occupying a small proportion of the scanned area. By reducing the FOV there is a reduction in distortion and an increase in in-plane resolution. The study described in this thesis is the first study which has assessed the diagnostic accuracy of a small FOV DWI technique with radical prostatectomy as a reference standard and the first which combined the technique with b-values greater than 1000 sec/mm^2 .

A prospective observation study comparing a small FOV technique against a more conventional large FOV sequence matched for b-values (b100, b1000 and b1500) and slice thickness was performed. Two radiologists blindly assessed 40 anonymised cases using DWI alone and with T2WI. Each DWI sequence assessed with a 2-week gap and in a random order to reduce recall bias. Readers recorded their findings on a reporting template which was identical to the histopathology reporting template to improve correlation.

The results concluded that the conventional DWI sequence was more sensitive, reliable and accurate, and the small FOV sequence more specific, with some of the differences reaching statistical significance. ADC of tumour was less than normal tissue for both sequences, but was significantly lower in small FOV than conventional DWI.

The diagnostic performance of both sequences was not as good as much of the published literature and the meta-analyses presented in this thesis, particularly the sensitivity. This is explained by the possible inclusion of much smaller tumours in the FODIP study, and potentially publication bias. Additionally, the assessment for tumour multifocality within the prostate resulted in the inclusion of all tumours rather than the index tumour, which is usually larger and/or more aggressive than the other tumours within the gland and therefore easier to detect and locate.

The quantitative results demonstrated a weak correlation in tumour ADC value and negative correlation with both DWI sequences. This may have been influenced by the large number of tumours which were situated in an area of post-biopsy haemorrhage. However, the weak correlation and significant overlap of values of ADC values in tumours of different grade means that the use of either sequence as a reliable non-invasive means of predicting tumour aggressiveness is not advised.

It is unlikely the small FOV technique will replace conventional DWI, but the use of high b-values will persist. The improved overall diagnostic performance of conventional imaging is not the only benefit of this sequence, the greater coverage of the pelvis allows assessment for pelvic lymphadenopathy and bony disease, as well as the reduced scan time, are further indications for the continued use of this sequence. The strength of the small FOV technique is its high specificity and thus its ability to correctly rule out significant prostate cancer. As the recent PROMIS trial has suggested there is a benefit in performing MRI before rather than after biopsy and this practice is being implemented at most hospitals in England. Thus a specific test, like small FOV DWI, in this scenario would non-invasively allow the more appropriate selection of men who do not have clinically significant prostate cancer and therefore do not require biopsy, which has benefit to the patient and reduces cost.

Definitions

ADC	Apparent Diffusion Coefficient	PZ	Peripheral Zone
AFMS	Anterior fibromuscular stroma	ROC	Receiver Operating Characteristic Curve
cDWI	Conventional DWI sequence	ROI	Region of Interest
CI	Confidence Interval	sDWI	Small FOV DWI sequence
CZ	Central Zone	SD	Standard Deviation
DCE	Dynamic Contrast Enhancement	SI	Signal Intensity
DRE	Digital Rectal Examination	sROC	Summary ROC Curve
DWI	Diffusion-weighted Imaging	T1WI	T1 Weighted Imaging
EPI	Echo Planar Imaging	T2WI	T2 Weighted Imaging
FOV	Field-of-View	TE	Echo Time
FOCUS	Field-of-view Optimized and Constrained Undistorted Single-Shot sequence	TNM	Tumour Node Metastasis staging system
IQR	Interquartile Range	TP	Transperineal biopsy
mpMRI	Multiparametric MRI	TRUS	Trans-rectal Ultrasound biopsy
MRI	Magnetic Resonance Imaging	TURP	Trans-urethral resection of prostate
NNUH	Norfolk and Norwich University Hospital	TZ	Transition Zone
PACS	Picture Archiving and Communication System		

References

1. McNeal JE. The zonal anatomy of the prostate. *Prostate*. 1981;2(1):35–49.
2. Fine SW, Reuter VE. Anatomy of the prostate revisited: Implications for prostate biopsy and zonal origins of prostate cancer. *Histopathology*. 2012;60(1):142–52.
3. Bhavsar A, Verma S. Anatomic imaging of the prostate. *Biomed Res Int*. 2014;2014:728539.
4. Lee CH, Akin-Olugbade O, Kirschenbaum A. Overview of Prostate Anatomy, Histology, and Pathology. *Endocrinol Metab Clin North Am*. 2011 Sep;40(3):565–75.
5. Ayala AG, Ro JY, Babaian R, Troncoso P, Grignon DJ. The prostatic capsule: does it exist? Its importance in the staging and treatment of prostatic carcinoma. Vol. 13, *American Journal of Surgical Pathology*. 1989. p. 21–7.
6. Myers RP. Detrusor apron, associated vascular plexus, and avascular plane: relevance to radical retropubic prostatectomy--anatomic and surgical commentary. *Urology*. 2002 Apr;59(4):472–9.
7. Kinugasa Y, Murakami G, Uchimoto K, Takenaka A, Yajima T, Sugihara K. Operating behind Denonvilliers' fascia for reliable preservation of urogenital autonomic nerves in total mesorectal excision: a histologic study using cadaveric specimens, including a surgical experiment using fresh cadaveric models. *Dis Colon Rectum*. 2006 Jul;49(7):1024–32.
8. Fuse H, Okumura A, Satomi S, Kazama T, Katayama T. Evaluation of seminal vesicle characteristics by ultrasonography before and after ejaculation. *Urol Int*. 1992;49(2):110–3.
9. National Collaborating Centre for Cancer. Prostate Cancer : diagnosis and treatment. Clinical guideline. *Natl Inst Heal Care Excell*. 2014;(January):1–480.
10. Prostate cancer statistics. Cancer Research UK [Internet]. [cited 2016 May 27]. Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer>.
11. Urological Cancers SSCRG. Mortality from Prostate Cancer. *Natl Cancer Intell Netw*. 2012;1–16.

12. Prostate cancer mortality. Cancer Research UK [Internet]. [cited 2018 February 16] . <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/mortality>.
13. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics*. 2008;64(1):10–9.
14. Prostate cancer incidence statistics. Cancer Research UK [Internet]. [cited 2018 February 16]. <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence>.
15. Ben-Shlomo Y, Evans S, Ibrahim F, Patel B, Anson K, Chinegwundoh F, Corbishley C, Dorling D, Thomas B, Gillatt D, Kirby R, Muir G, Nargund V, Popert R, Metcalfe C, Persad R. The Risk of Prostate Cancer amongst Black Men in the United Kingdom: The PROCESS Cohort Study. *Eur Urol*. 2008;53(1):99–105.
16. Shand RL, Gelmann EP. Molecular biology of prostate-cancer pathogenesis. *Curr Opin Urol*. 2006;16:123–31.
17. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, Garofalo A, Gulati R, Carreira S, Eeles R, Elemento O, Rubin MA, Robinson D, Lonigro R, Hussain M, Chinnaiyan A, Vinson J, Filipenko J, Garraway L, Taplin M-E, AlDubayan S, Han GC, Beightol M, Morrissey C, Nghiem B, Cheng HH, Montgomery B, Walsh T, Casadei S, Berger M, Zhang L, Zehir A, Vijai J, Scher HI, Sawyers C, Schultz N, Kantoff PW, Solit D, Robson M, Van Allen EM, Offit K, de Bono J, Nelson PS. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med*. 2016;375(5):443–53.
18. Castro E, Eeles R. The role of BRCA1 and BRCA2 in prostate cancer. *Asian J Androl*. 2012;14(3):409–14.
19. Klap J, Schmid M, Loughlin KR. The relationship between total testosterone levels and prostate cancer: A review of the continuing controversy. *J Urol*. 2015;193(2):403–13.
20. Brinker DA, Potter SR, Epstein JI. Ductal adenocarcinoma of the prostate diagnosed on needle biopsy: correlation with clinical and radical prostatectomy findings and progression. *Am J Surg Pathol*. 1999;23(12):1471–9.
21. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO Classification of

- Tumours of the Urinary System and Male Genital Organs—Part B: Prostate and Bladder Tumours. *Eur Urol.* 2016;70(1):106–19.
22. Zhou M. Intraductal carcinoma of the prostate: the whole story. *Pathology.* 2013 Oct;45(6):533–9.
23. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother reports.* 1966 Mar;50(3):125–8.
24. D’Amico A V, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ, Wein A. Biochemical Outcome After Radical Prostatectomy , External Beam Radiation Therapy , or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer. *JAMA.* 1998;280(11):969–74.
25. Gleason D. Histologic grading of prostate cancer: a perspective. *Hum Pathol.* 1992 Mar;23(3):273–9.
26. Epstein JI, Allsbrook WC, Amin MB. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *J Am Pathol.* 2005;29(9):1228–42.
27. Helpap B, Egevad L. The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens. *Virchows Arch.* 2006;449:622–7.
28. Billis A, Guimaraes MS, Freitas LLL, Meirelles L, Magna LA, Ferreira U. The Impact of the 2005 International Society of Urological Pathology Consensus Conference on Standard Gleason Grading of Prostatic Carcinoma in Needle Biopsies. *J Urol.* 2008;180(August):548–53.
29. Eggener SE, Scardino PT, Walsh PC, Han M, Partin AW, Trock BJ, Feng Z, Wood DP, Eastham JA, Rabah DM, Kattan MW, Yu C, Klein EA. Predicting 15-year prostate cancer specific mortality after radical prostatectomy. *J Urol.* 2011;185(3):869–75.
30. Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping : data based on the modified Gleason scoring system. *BJU Int.* 2013 May;111(5):753–60.
31. Steinberg DM, Sauvageot J, Piantadosi S, Epstein JI. Correlation of prostate needle biopsy and radical prostatectomy Gleason grade in academic and community settings. *Am J Surg Pathol.* 1997 May;21(5):566–76.

32. Epstein JI, Chan DW, Sokoll LJ, Walsh PC, Cox JL, Rittenhouse H, Wolfert R, Carter HB. Nonpalpable stage T1c prostate cancer: prediction of insignificant disease using free/total prostate specific antigen levels and needle biopsy findings. *J Urol*. 1998;160(6):2407–11.
33. Bastian PJ, Mangold LA, Epstein JI, Partin AW, Ph D. Characteristics of Insignificant Clinical T1c Prostate Tumors. A contemporary analysis. *Cancer*. 2004;101(9):2001–5.
34. Johnstone PAS, Rossi PJ, Jani AB, Master V. “ Insignificant ” prostate cancer on biopsy : pathologic results from subsequent radical prostatectomy. *Prostate Cancer Prostatic Dis*. 2007;10(3):237–41.
35. Arora R, Koch MO, Eble JN, Ulbright TM, Li L, Ph D, Cheng L. Heterogeneity of Gleason Grade in Multifocal Adenocarcinoma of the Prostate. 2004;(April):2362–6.
36. Wise AM, Stamey TA, McNeal JE, Clayton JL. Morphologic and clinical significance of multifocal prostate cancers in radical prostatectomy specimens. *Urology*. 2002;60(2):264–9.
37. Le JD, Tan N, Shkoliar E, Lu DY, Kwan L, Marks LS, Huang J, Margolis DJAA, Raman SS, Reiter RE. Multifocality and Prostate Cancer Detection by Multiparametric Magnetic Resonance Imaging: Correlation with Whole-mount Histopathology. *Eur Urol*. 2014 Sep 22;67(3):569–76.
38. Mehra R, Han B, Tomlins SA, Wang L, Menon A, Wasco MJ, Shen R, Montie JE, Chinnaiyan AM, Shah RB. Heterogeneity of TMPRSS2 Gene Rearrangements in Multifocal Prostate Adenocarcinoma : Molecular Evidence for an Independent Group of Diseases. *Cancer Res*. 2007;67(17):7991–6.
39. Boutros PC, Fraser M, Harding NJ, De Borja R, Trudel D, Lalonde E, Meng A, Hennings-Yeomans PH, McPherson A, Sabelnykova VY, Zia A, Fox NS, Livingstone J, Shiah YJ, Wang J, Beck TA, Have CL, Chong T, Sam M, Johns J, Timms L, Buchner N, Wong A, Watson JD, Simmons TT, P’ng C, Zafarana G, Nguyen F, Luo X, Chu KC, Prokopec SD, Sykes J, Pra AD, Berlin A, Brown A, Chan-Seng-Yue MA, Yousif F, Denroche RE, Chong LC, Chen GM, Jung E, Fung C, Starmans MHW, Chen H, Govind SK, Hawley J, D’Costa A, Pintilie M, Waggott D, Hach F, Lambin P, Muthuswamy LB, Cooper C, Eeles R, Neal D, Tetu B, Sahinalp C, Stein LD, Fleshner N, Shah SP, Collins CC, Hudson TJ, McPherson JD, Van Der Kwast T, Bristow RG. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet*.

- 2015;47(7):736–45.
40. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, Nickerson E, Chae SS, Boysen G, Auclair D, Onofrio RC, Park K, Kitabayashi N, MacDonald TY, Sheikh K, Vuong T, Guiducci C, Cibulskis K, Sivachenko A, Carter SL, Saksena G, Voet D, Hussain WM, Ramos AH, Winckler W, Redman MC, Ardlie K, Tewari AK, Mosquera JM, Rupp N, Wild PJ, Moch H, Morrissey C, Nelson PS, Kantoff PW, Gabriel SB, Golub TR, Meyerson M, Lander ES, Getz G, Rubin MA, Garraway LA. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet.* 2012;44(6):685–9.
 41. Lindberg J, Klevebring D, Liu W, Neiman M, Xu J, Wiklund P, Wiklund F, Mills IG, Egevad L, Grönberg H. Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins. *Eur Urol.* 2013;63(2):347–53.
 42. Andreoiu M, Cheng L. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. *Hum Pathol.* 2010;41(6):781–93.
 43. van der Kwast TH, Amin MB, Billis A, Epstein JI, Griffiths D, Humphrey PA, Montironi R, Wheeler TM, Srigley JR, Egevad L, Delahunt B, ISUP Prostate Cancer Group. International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 2: T2 substaging and prostate cancer volume. *Mod Pathol.* 2011 Jan 3;24(1):16–25.
 44. Huang CC, Deng F-M, Kong MX, Ren Q, Melamed J, Zhou M. Re-evaluating the concept of “dominant/index tumor nodule” in multifocal prostate cancer. *Virchows Arch.* 2014;464(5):589–94.
 45. American Joint Committee on Cancer: Prostate cancer staging. 7th Edition. 2009.
 46. Ilyin SE, Belkowski SM, Plata-salama CR. Biomarker discovery and validation : technologies and integrative approaches. *Trends Biotechnol.* 2004;22(8):411–6.
 47. Lilja H, Oldbring J, Rannevik G, Laurell C. Seminal Vesicle-secreted Proteins and Their Reactions during Gelation and Liquefaction of Human Semen. *J Clin investogation.* 1987;80(August):281–5.
 48. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM,

- Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA. Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤ 4.0 ng per Milliliter. *N Engl J Med*. 2004;350:2239–46.
49. Carvalhal GF, Daudi SN, Kan D, Mondo D, Roehl KA, Loeb S, Catalona WJ. Correlation Between Serum Prostate-specific Antigen and Cancer Volume in Prostate Glands of Different Sizes. *Urology*. 2010;76(5):1072–6.
50. Kato RB, Srougi V, Salvadori FA, Paulo P. Pretreatment tumor volume estimation based on total serum PSA in patients with localized prostate cancer. *Clinics*. 2008;63(6):759–62.
51. Holmstrom B, Johansson M, Bergh A, Stenman U, Hallmans G, Stattin P. Prostate specific antigen for early detection of prostate. *Bmj*. 2009;339:b3537.
52. Schroder FH, Hugosson J, Roobol MJ, Tammela TLJ, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Berenguer A, Määttänen L, Bangma CH, Aus G, Villers A, Rebillard X, Kwast T Van Der, Blijenberg BG, Moss SM, Koning HJ De, Auvinen A, Investigators E. Screening and Prostate-Cancer Mortality in a Randomized European Study. *N Engl J Med*. 2009;360:1320–8.
53. Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, Feuer E, Koning H De. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening : Importance of Methods and Context. *J Natl Cancer Inst*. 2009;101(6):374–83.
54. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Kwast T Van Der, Mason M, Matveev V, Wiegel T, Zattoni F, Mottet N. EAU Guidelines on Prostate Cancer . Part 1 : Screening , Diagnosis , and Local Treatment with Curative Intent — Update 2013. *Eur Urol*. 2014;65(1):124–37.
55. Moyer VA. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012 Jul 17;157(2):120–34.
56. Bonekamp D, Jacobs MA, El-Khouli R, Stoianovici D, Macura KJ. Advancements in MR Imaging of the Prostate: From Diagnosis to Interventions. *Radiographics*. 2011 Jan;31(3):677–703.
57. Welch HG, Fisher ES, Gottlieb DJ, Barry MJ. Detection of Prostate Cancer via Biopsy in the Medicare – SEER Population During the PSA Era. *J Natl Cancer Inst*. 2007;19(18):1395–400.

58. Ahmed HU, Bosaily AE, Brown LC, Gabe R, Kaplan R, Parmar MK, Collaco-moraes Y. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. 2017;389:815–22.
59. Bott SRJ, Young MPA, Kellett MJ, Parkinson MC. Anterior prostate cancer : is it more difficult to diagnose ? *BJU Int*. 2002;89:886–9.
60. Williamson DA, Barrett LK, Rogers BA, Freeman JT, Hadway P, Paterson DL. Infectious Complications Following Transrectal Ultrasound – Guided Prostate Biopsy : New Challenges in the Era of Multidrug-Resistant Escherichia coli. *Clin Infect Dis*. 2013;57(2):267–74.
61. Batura D, Rao GG. The national burden of infections after prostate biopsy in England and Wales : a wake-up call for better prevention. *J Antimicrob Chemotherapy*. 2013;68:247–9.
62. Shen P, Zhu Y, Wei W, Li Y, Yang J, Li Y, Li D. The results of transperineal versus transrectal prostate biopsy : a systematic review and meta-analysis. *Asian J Androl*. 2012;14(2):310–5.
63. Zlotta AR, Egawa S, Pushkar D, Govorov A, Kimura T, Kido M, Takahashi H, Kuk C, Kovylyina M, Aldaoud N, Fleshner N, Finelli A, Klotz L, Sykes J, Lockwood G. Prevalence of Prostate Cancer on Autopsy : Cross-Sectional Study on Unscreened Caucasian and Asian Men. *J Natl Cancer Inst*. 2013;105(14):1050–8.
64. European Association of Urology - Prostate cancer guidelines. 2016.
65. Haglind E, Carlsson S, Stranne J, Wallerstedt A, Wildera U, Thorsteinsdottir T, Lagerkvist M, Damber J, Bjartell A, Hugosson J, Wiklund P, Steineck G. Urinary Incontinence and Erectile Dysfunction After Robotic Versus Open Radical Prostatectomy : A Prospective . *Eur Urol*. 2015;68(2):216–25.
66. Ash D, Flynn A, Battermann J, Reijke T De, Lavagnini P, Blank L. ESTRO / EAU / EORTC recommendations on permanent seed implantation for localized prostate cancer. *Radiother Oncol*. 2000;57(3):315–21.
67. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, Margolis D, Schnall MD, Shtern F, Tempany CM, Thoeny HC, Verma S. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol*. 2016;69(1):16–40.
68. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, Rouviere O, Logager V, Fütterer JJ. ESUR prostate MR guidelines 2012. *Eur Radiol*. 2012 Apr;22(4):746–57.

69. Barrett T, Vargas HA, Akin O, Goldman DA, Hricak H. Value of the hemorrhage exclusion sign on T1-weighted prostate MR images for the detection of prostate cancer. *Radiology*. 2012;263(3):751–7.
70. White S, Hricak H, Forstner R, Kurhanewicz J, Vigneron DB, Zaloudek CJ, Weiss JM, Narayan P, Carroll PR. Prostate cancer: effect of postbiopsy hemorrhage on interpretation of MR images. *Radiology*. 1995;195(2):385–90.
71. Park KK, Lee SH, Lim BJ, Kim JH, Chung BH, Dickinson L, Ahmed HU, Moore C, Freeman A, Kirkham A, Allen C, Emberton M, Park KK, Lee SH, Lim BJ, Kim JH, Chung BH. The effects of the period between biopsy and diffusion-weighted magnetic resonance imaging on cancer staging in localized prostate cancer. *BJU Int*. 2010 Oct;106(8):1148–51.
72. Rosenkrantz AB, Mussi TC, Hindman N, Lim RP, Kong MX, Babb JS, Melamed J, Taneja SS. Impact of delay after biopsy and post-biopsy haemorrhage on prostate cancer tumour detection using multi-parametric MRI: A multi-reader study. *Clin Radiol*. 2012;67(12):e83--e90.
73. Rosenkrantz AB, Kopec M, Kong X, Melamed J, Dakwar G, Babb JS, Taouli B. Prostate cancer vs. post-biopsy hemorrhage: Diagnosis with T2- and diffusion-weighted imaging. *J Magn Reson Imaging*. 2010;31(6):1387–94.
74. Gibbs P, Tozer DJ, Liney GP, Turnbull LW. Comparison of quantitative T-2 mapping and diffusion-weighted imaging in the normal and pathologic prostate. *Magn Reson Med*. 2001;46(6):1054–8.
75. Murphy G, Haider M, Ghai S, Sreeharsha B. The expanding role of MRI in prostate cancer. *AJR Am J Roentgenol*. 2013 Dec;201(6):1229–38.
76. Neil JJ. Diffusion imaging concepts for clinicians. *J Magn Reson Imaging*. 2008;27(1):1–7.
77. Sugahara T, Korogi Y, Kochi M, Ikushima I, Shigematu Y, Hirai T, Okuda T, Liang L, Ge Y, Komohara Y, Ushio Y, Takahashi M. Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas. *J Magn Reson Imaging*. 1999;9(1):53–60.
78. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: Applications and challenges in oncology. *Am J Roentgenol*. 2007;188(6):1622–35.

79. Stejskal EO, Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J Chem Phys.* 1965;42(1):288–92.
80. Hambrock T, Somford DM, Huisman HJ, Van Oort IM, Witjes JA, Hulsbergen-Van De Kaa CA, Scheenen T, Barentsz JO. Relationship between apparent diffusion coefficients at 3.0-T mr imaging and gleason grade in peripheral zone prostate cancer. *Radiology.* 2011;259(2):453–61.
81. Rosenkrantz AB, Hindman N, Lim RP, Das K, Babb JS, Mussi TC, Taneja SS. Diffusion-weighted imaging of the prostate: Comparison of b1000 and b2000 image sets for index lesion detection. *J Magn Reson Imaging.* 2013 Sep;38(3):694–700.
82. Rosenkrantz AB, Kong X, Niver BE, Berkman DS, Melamed J, Babb JS, Taneja SS. Prostate Cancer: Comparison of Tumor Visibility on Trace Diffusion-Weighted Images and the Apparent Diffusion Coefficient Map. *Am J Roentgenol.* 2011;196(1):123–9.
83. Ueno Y, Kitajima K, Sugimura K, Kawakami F, Miyake H, Obara M, Takahashi S. Ultra-high b-value diffusion-weighted MRI for the detection of prostate cancer with 3-T MRI. *J Magn Reson Imaging.* 2013;38(1):154–60.
84. Tamada T, Kanomata N, Sone T, Jo Y, Miyaji Y, Higashi H, Yamamoto A, Ito K. High b value (2,000 s/mm²) diffusion-weighted magnetic resonance imaging in prostate cancer at 3 Tesla: comparison with 1,000 s/mm² for tumor conspicuity and discrimination of aggressiveness. *PLoS One.* 2014 Jan;9(5):e96619.
85. Ohgiya Y, Suyama J, Seino N, Hashizume T, Kawahara M, Sai S, Saiki M, Munechika J, Hirose M, Gokan T. Diagnostic accuracy of ultra-high-b-value 3.0-T diffusion-weighted MR imaging for detection of prostate cancer. *Clin Imaging.* 2012 Oct;36(5):526–31.
86. Katahira K, Takahara T, Kwee TC, Oda S, Suzuki Y, Morishita S, Kitani K, Hamada Y, Kitaoka M, Yamashita Y. Ultra-high-b-value diffusion-weighted MR imaging for the detection of prostate cancer: Evaluation in 201 cases with histopathological correlation. *Eur Radiol.* 2011;21(1):188–96.
87. Metens T, Miranda D, Absil J, Matos C. What is the optimal b value in diffusion-weighted MR imaging to depict prostate cancer at 3T? *Eur Radiol.* 2012;22(3):703–9.
88. Wang X, Qian Y, Liu B, Cao L, Fan Y, Zhang JJ, Yu Y. High-b-value diffusion-weighted MRI for

- the detection of prostate cancer at 3 T. *Clin Radiol*. 2014 Nov;69(11):1165–70.
89. Wetter A, Nensa F, Lipponer C, Guberina N, Olbricht T, Schenck M, Schlosser TW, Gratz M, Lauenstein TC. High and ultra-high b-value diffusion-weighted imaging in prostate cancer: a quantitative analysis. *Acta Radiol*. 2014 Aug 28;56(8):1009–15.
90. 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 s/mm². *AJR Am J Roentgenol*. 2010 Jan;194(1):W33-7.
91. Koo JH, Kim CK, Choi D, Park BK, Kwon GY, Kim B. Diffusion-weighted magnetic resonance imaging for the evaluation of prostate cancer: Optimal B value at 3T. *Korean J Radiol*. 2013;14(1):61–9.
92. Kitajima K, Takahashi S, Ueno Y, Yoshikawa T, Ohno Y, Obara M, Miyake H, Fujisawa M, Sugimura K. Clinical utility of apparent diffusion coefficient values obtained using high b-value when diagnosing prostate cancer using 3 tesla MRI: comparison between ultra-high b-value (2000 s/mm²) and standard high b-value (1000 s/mm²). *J Magn Reson Imaging*. 2012 Jul;36(1):198–205.
93. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988 Aug 1;168(2):497–505.
94. Thoermer G, Otto J, Reiss-Zimmermann M, Seiwerts M, Moche M, Garnov N, Franz T, Do M, Stolzenburg J-U, Horn L-C, Kahn T, Busse H. Diagnostic value of ADC in patients with prostate cancer: influence of the choice of b values. *Eur Radiol*. 2012;22(8):1820–8.
95. Hosseinzadeh K, Schwarz SD. Endorectal diffusion-weighted imaging in prostate cancer to differentiate malignant and benign peripheral zone tissue. *J Magn Reson Imaging*. 2004;20(4):654–61.
96. Kim CK, Park BK, Han JJ, Kang TW, Lee HM. Diffusion-weighted imaging of the prostate at 3 T for differentiation of malignant and benign tissue in transition and peripheral zones: preliminary results. *J Comput Assist Tomogr*. 2007;31(3):449–54.
97. Mazaheri Y, Shukla-Dave A, Hricak H, Fine SW, Zhang J, Inurrigarro G, Moskowitz CS, Ishill NM, Reuter VE, Touijer K, Zakian KL, Koutcher JA. Prostate cancer: Identification with

- combined diffusion-weighted MR imaging and 3D (HMR)-H-1 spectroscopic imaging correlation with pathologic findings. *Radiology*. 2008;246(2):480–8.
98. Tamada T, Sone T, Jo Y, Toshimitsu S, Yamashita T, Yamamoto A, Tanimoto D, Ito K. Apparent diffusion coefficient values in peripheral and transition zones of the prostate: comparison between normal and malignant prostatic tissues and correlation with histologic grade. *J Magn Reson Imaging*. 2008 Sep;28(3):720–6.
99. Kim CK, Park BK, Lee HM, Kwon GY. Value of diffusion-weighted imaging for the prediction of prostate cancer location at 3T using a phased-array coil: preliminary results. *Invest Radiol*. 2007;42(12):842–7.
100. Jie C, Rongbo L, Ping T, Chen J, Liu R, Tan P. The value of diffusion-weighted imaging in the detection of prostate cancer: a meta-analysis. *Eur Radiol*. 2014 Aug;24(8):1929–41.
101. Wu L-M, Xu J-R, Ye Y-Q, Lu Q, Hu J-N. The clinical value of diffusion-weighted imaging in combination with T2-weighted imaging in diagnosing prostate carcinoma: a systematic review and meta-analysis. *AJR Am J Roentgenol*. 2012 Jul;199(1):103–10.
102. Rosenkrantz AB, Ream JM, Nolan P, Rusinek H, Deng FM, Taneja SS. Prostate cancer: Utility of whole-lesion apparent diffusion coefficient metrics for prediction of biochemical recurrence after radical prostatectomy. *Am J Roentgenol*. 2015;205(6):1208–14.
103. Donati OF, Mazaheri Y, Afaq A, Vargas HA, Zheng J, Moskowitz CS, Hricak H, Akin O. Prostate cancer aggressiveness: assessment with whole-lesion histogram analysis of the apparent diffusion coefficient. *Radiology*. 2014 Apr;271(1):143–52.
104. Vargas HA, Akin O, Franiel T, Mazaheri Y, Zheng J, Moskowitz C, Udo K, Eastham J, Hricak H. Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: tumor detection and assessment of aggressiveness. *Radiology*. 2011;259(3):775–84.
105. Thormer G, Otto J, Horn L-CC, Garnov N, Minh D, Franz T, Stolzenburg J-UU, Moche M, Kahn T, Busse H, Thormer G, Otto J, Horn L-CC, Garnov N, Do M, Franz T, Stolzenburg J-UU, Moche M, Kahn T, Busse H, Thormer G, Otto J, Horn L-CC, Garnov N, Minh D, Franz T, Stolzenburg J-UU, Moche M, Kahn T, Busse H. Non-invasive estimation of prostate cancer aggressiveness using diffusion-weighted MRI and 3D proton MR spectroscopy at 3.0 T. *Acta radiol*. 2015;56(1):121–8.

106. Nagarajan R, Margolis D, Raman S, Sarma MK, Sheng K, King CR, Verma G, Sayre J, Reiter RE, Thomas MA. MR spectroscopic imaging and diffusion-weighted imaging of prostate cancer with Gleason scores. *J Magn Reson Imaging*. 2012;36(3):697–703.
107. Peng Y, Jiang Y, Antic T, Sethi I, Schmid-Tannwald C, Eggener S, Oto A. Apparent Diffusion Coefficient for Prostate Cancer Imaging: Impact of b Values. *Am J Roentgenol*. 2014;202(3):W247–53.
108. Wu X, Reinikainen P, Vanhanen A, Kapanen M, Vierikko T, Ryymin P, Hyödynmaa S, Kellokumpu-Lehtinen P-L. Correlation between apparent diffusion coefficient value on diffusion-weighted MR imaging and Gleason score in prostate cancer. *Diagn Interv Imaging*. 2017;98(1):63–71.
109. Medved M, Soylu-Boy FN, Karademir I, Sethi I, Yousuf A, Karczmar GS, Oto A. High-resolution diffusion-weighted imaging of the prostate. *AJR Am J Roentgenol*. 2014 Jul 20;203(1):85–90.
110. Korn N, Kurhanewicz J, Banerjee S, Starobinets O, Saritas E, Noworolski S. Reduced-FOV excitation decreases susceptibility artifact in diffusion-weighted MRI with endorectal coil for prostate cancer detection. *Magn Reson Imaging*. 2015;33(1):56–62.
111. Saritas EU, Cunningham CH, Lee JH, Han ET, Nishimura DG. DWI of the spinal cord with reduced FOV single-shot EPI. *Magn Reson Med*. 2008;60:468–73.
112. Zaharchuk G, Saritas EU, Andre JB, Chin CT, Rosenberg J, Brosnan TJ, Shankaranarayan A, Nishimura DG, Fischbein NJ. Reduced field-of-view diffusion imaging of the human spinal cord: comparison with conventional single-shot echo-planar imaging. *AJNR Am J Neuroradiol*. 2011 May;32(May):813–20.
113. Riffel P, Michaely HJ, Morelli JN, Pfeuffer J, Attenberger UI, Schoenberg SO, Haneder S. Zoomed EPI-DWI of the pancreas using two-dimensional spatially-selective radiofrequency excitation pulses. *PLoS One*. 2014 Jan;9(3):e89468.
114. Dong H, Li Y, Li H, Wang B, Hu B. Study of the reduced field-of-view diffusion-weighted imaging of the breast. *Clin Breast Cancer*. 2014;14(4):265–71.
115. Feng Z, Min X, Sah VK, Li L, Cai J, Deng M, Wang L. Comparison of field-of-view (FOV) optimized and constrained undistorted single shot (FOCUS) with conventional DWI for the

- evaluation of prostate cancer. *Clin Imaging*. 2015;39(5):851–5.
116. Rosenkrantz AB, Chandarana H, Pfeuffer J, Triolo MJ, Shaikh MB, Mossa DJ, Geppert C. Zoomed echo-planar imaging using parallel transmission: impact on image quality of diffusion-weighted imaging of the prostate at 3T. *Abdom Imaging*. 2014 Jun 25;
117. Thierfelder KM, Sommer WH, Dietrich O, Meinel FG, Theisen D, Paprottka PM, Strobl FF, Pfeuffer J, Reiser MF, Nikolaou K. Parallel-transmit-accelerated spatially-selective excitation MRI for reduced-FOV diffusion-weighted-imaging of the pancreas. *Eur J Radiol*. 2014 Oct;83(10):1709–14.
118. Attenberger UI, Rathmann N, Sertdemir M, Riffel P, Weidner A, Kannengiesser S, Morelli JN, Schoenberg SO, Hausmann D. Small Field-of-view single-shot EPI-DWI of the prostate: Evaluation of spatially-tailored two-dimensional radiofrequency excitation pulses. *Z Med Phys*. 2016;26(2):168–76.
119. Brendle C, Martirosian P, Schwenzer NF, Kaufmann S, Kruck S, Kramer U, Notohamiprodjo M, Nikolaou K, Schraml C. Diffusion-weighted imaging in the assessment of prostate cancer: Comparison of zoomed imaging and conventional technique. *Eur J Radiol*. 2016;85(5):893–900.
120. Wibner AG, Burger IA, Sala E, Hricak H, Weber WA, Vargas HA. Molecular Imaging of Prostate Cancer. *Radiographics*. 2016;36(1):142–59.
121. Eiber M, Nekolla SG, Maurer T, Weirich G, Wester HJ, Schwaiger M. ⁶⁸Ga-PSMA PET/MR with multimodality image analysis for primary prostate cancer. *Abdom Imaging*. 2015;40(6):1769–71.
122. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, Group Q-2. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–36.
123. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, Meulen JHP van der, Bossuyt PMM. Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests. *JAMA*. 1999 Sep 15;282(11):1061–6.
124. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW. Towards Complete and Accurate Reporting of Studies of Diagnostic

- Accuracy: The STARD Initiative. *Radiology*. 2003 Jan 1;226(1):24–8.
125. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N, Bossuyt PMM. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology*. 2015 Mar;274(3):781–9.
126. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology*. 2015 Dec;277(3):826–32.
127. Samaratunga H, Montironi R, True L, Epstein JI, Griffiths DF, Humphrey P a, van der Kwast TH, Wheeler TM, Srigley JR, Delahunt B, Egevad L, Amin MB, Billis A, Epstein JI, Griffiths DF, Humphrey P a, Montironi R, Wheeler TM, Srigley JR, Egevad L, Delahunt B. International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 1: specimen handling. *Mod Pathol*. 2011 Jan;24(1):6–15.
128. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.
129. Mulrow CD. Rationale for systematic reviews. *Br Med J*. 1994;309(6954):597–9.
130. Crighton EJ, Mamdani MM, Upshur REG. *BMC Health Services Research*. *BMC Health Serv Res*. 2009;8(1410):1–8.
131. Jin G, Su DK, Luo N Bin, Liu LD, Zhu X, Huang XY. Meta-analysis of diffusion-weighted magnetic resonance imaging in detecting prostate cancer. *J Comput Assist Tomogr*. 2013;37(2):195–202.
132. Mazzucchelli R, Scarpelli M, Cheng L, Lopez-Beltran A, Galosi AB, Kirkali Z, Montironi R. Pathology of prostate cancer and focal therapy ('male lumpectomy'). *Anticancer Res*. 2009;29(12):5155–61.
133. Turkbey B, Pinto PA, Mani H, Bernardo M, Pang Y, McKinney YL, Khurana K, Ravizzini GC, Albert PS, Merino MJ, Choyke PL. Prostate cancer: value of multiparametric MR imaging at 3 T for detection--histopathologic correlation. *Radiology*. 2010 Apr;255(1):89–99.

134. Rosenkrantz AB, Deng FM, Kim S, Lim RP, Hindman N, Mussi TC, Spieler B, Oaks J, Babb JS, Melamed J, Taneja SS. Prostate cancer: Multiparametric mri for index lesion localization - A multiple-reader study. *Am J Roentgenol*. 2012;199(4):830–7.
135. Kajihara H, Hayashida Y, Murakami R, Katahira K, Nishimura R, Hamada Y, Kitani K, Kitaoka M, Suzuki Y, Kitajima M, Hirai T, Morishita S, Awai K, Yamashita Y. Usefulness of Diffusion-Weighted Imaging in the Localization of Prostate Cancer. *Int J Radiat Oncol Biol Phys*. 2009;74(2):399–403.
136. Rosenkrantz AB, Chandarana H, Hindman N, Deng F-MM, Babb JS, Taneja SS, Geppert C. Computed diffusion-weighted imaging of the prostate at 3 T: Impact on image quality and tumour detection. *Eur Radiol*. 2013;23(11):3170–7.
137. Turkbey B, Merino MJ, Gallardo EC, Shah V, Aras O, Bernardo M, Mena E, Daar D, Rastinehad AR, Linehan WM, Wood BJ, Pinto PA, Choyke PL. Comparison of endorectal coil and nonendorectal coil T2W and diffusion-weighted MRI at 3 Tesla for localizing prostate cancer: Correlation with whole-mount histopathology. *J Magn Reson Imaging*. 2014;39(6):1443–8.
138. Kitajima K, Kaji Y, Fukabori Y, Yoshida K, Suganuma N, Sugimura K. Prostate cancer detection with 3 T MRI: Comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-weighted imaging. *J Magn Reson Imaging*. 2010;31(3):625–31.
139. Ueno Y, Takahashi S, Ohno Y, Kitajima K, Yui M, Kassai Y, Kawakami F, Miyake H, Sugimura K. Computed diffusion-weighted MRI for prostate cancer detection: The influence of the combinations of b-values. *Br J Radiol*. 2015;88(1048).
140. Haider MA, Van Der Kwast TH, Tanguay J, Evans AJ, Hashmi AT, Lockwood G, Trachtenberg J. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. *Am J Roentgenol*. 2007;189(2):323–8.
141. Isebaert S, Van den Bergh L, Haustermans K, Joniau S, Lerut E, De Wever L, De Keyzer F, Budiharto T, Slagmolen P, Van Poppel H, Oyen R. Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology. *J Magn Reson Imaging*. 2013 Jun;37(6):1392–401.
142. Lim HK, Kim JK, Kim K a., Cho KS. Prostate cancer: Apparent diffusion coefficient map with

- T2-weighted images for detection - A multireader study. *Radiology*. 2009;250(1):145–51.
143. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Stat Med*. 1993;12(14):1293–316.
144. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ Br Med J*. 2003;327(7414):557–60.
145. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. *Statistical Methods of Meta-Analysis*. *Korean J Radiol*. 2015;16(6):1188–96.
146. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–58.
147. Chen M, Dang H-D, Wang J-Y, Zhou C, Li S-Y, Wang W-C, Zhao W-F, Yang Z-H, Zhong C-Y, Li G-Z. Prostate cancer detection: comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined. *Acta radiol*. 2008;49(5):602–10.
148. Girometti R, Bazzocchi M, Como G, Brondani G, Del Pin M, Frea B, Martinez G, Zuiani C. Negative predictive value for cancer in patients with “gray-Zone” PSA level and prior negative biopsy: Preliminary results with multiparametric 3.0 tesla MR. *J Magn Reson Imaging*. 2012;36:943–50.
149. Kumar V, Jagannathan NR, Kumar R, Thulkar S, Gupta SD, Dwivedi SN, Hemal AK, Gupta NP. Apparent diffusion coefficient of the prostate in men prior to biopsy: determination of a cut-off value to predict malignancy of the peripheral zone. *NMR Biomed*. 2007;20(5):505–11.
150. Rosenkrantz ABB, Khalef V, Xu W, Babb JSS, Taneja SSS, Doshi AMM. Does normalisation improve the diagnostic performance of apparent diffusion coefficient values for prostate cancer assessment? A blinded independent-observer evaluation. *Clin Radiol*. 2015;70(9):1032–7.
151. Vilanova JC, Barceló-Vidal C, Comet J, Boada M, Barceló J, Ferrer J, Albanell J. Usefulness of prebiopsy multifunctional and morphologic MRI combined with free-to-total prostate-

- specific antigen ratio in the detection of prostate cancer. *Am J Roentgenol*. 2011;196(6):715–22.
152. Peng Y, Jiang Y, Yang C, Brown JB, Antic T, Sethi I, Schmid-Tannwald C, Giger ML, Eggener SE, Oto A. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score--a computer-aided diagnosis development study. *Radiology*. 2013 Jun;267(3):787–96.
153. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg*. 2005;79(1):16–20.
154. Handbook for DTA Reviews | Cochrane Screening and Diagnostic Tests [Internet]. [cited 2017 Sep 15]. Available from: <http://methods.cochrane.org/sdt/handbook-dta-reviews>
155. Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31.
156. Kim CK, Park BK, Kim B. Diffusion-weighted MRI at 3 T for the evaluation of prostate cancer. *Am J Roentgenol*. 2010;194(6):1461–9.
157. Kitajima K, Takahashi S, Ueno Y, Miyake H, Fujisawa M, Kawakami F, Sugimura K. Do apparent diffusion coefficient (ADC) values obtained using high b-values with a 3-T MRI correlate better than a transrectal ultrasound (TRUS)-guided biopsy with true Gleason scores obtained from radical prostatectomy specimens for patients with prostat. *Eur J Radiol*. 2013 Aug;82(8):1219–26.
158. Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, Richie JP, deKernion JB, Walsh PC, Scardino PT, Lange PH, Subong EN, Parson RE, Gasior GH, Loveland KG, Southwick PC. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. *JAMA*. 1998 May 20;279(19):1542–7.
159. Ahmed HU, Kirkham A, Arya M, Illing R, Freeman A, Allen C, Emberton M. Is it time to consider a role for MRI before prostate biopsy? *Nat Rev Clin Oncol*. 2009;6(4):197–206.
160. Noguchi M, Stamey TA, McNeal JE, NOLLEY R. Prognostic Factors for Multifocal Prostate Cancer in Radical Prostatectomy Specimens: Lack of Significance of Secondary Cancers. *J Urol*. 2003;170(2):459–63.

161. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011.
162. National Institute for Health Research. PROSPERO [Internet]. [cited 2017 May 29]. Available from: <https://www.crd.york.ac.uk/PROSPERO/>
163. Hoeks CMA, Hambroek T, Yakar D, Hulsbergen-van de Kaa C a, Feuth T, Witjes JA, Fütterer JJ, Barentsz JO, de Kaa CAH, Feuth T, Witjes JA, Fütterer JJ, Barentsz JO. Transition Zone Prostate Cancer: Detection and Localization with 3-T Multiparametric MR Imaging. *Radiology*. 2013;266(1):207–17.
164. Jung S Il, Donati OF, Vargas H a, Goldman D, Hricak H, Akin O. Transition Zone Prostate Cancer: Incremental Value of Diffusion-weighted Endorectal MR Imaging in Tumor Detection and Assessment of Aggressiveness. *Radiology*. 2013;269(2):493–503.
165. Kim JY, Kim SH, Kim YH, Lee HJ, Kim MJ, Choi MS. Low-risk prostate cancer: The accuracy of multiparametric MR imaging for detection. *Radiology*. 2014;271(2):435–44.
166. Tanimoto A, Nakashima J, Kohno H, Shinmoto H, Kuribayashi S. Prostate cancer screening: the clinical value of diffusion-weighted imaging and dynamic MR imaging in combination with T2-weighted imaging. *J Magn Reson Imaging*. 2007;25(1):146–52.
167. Petrillo A, Fusco R, Setola S V., Ronza FM, Granata V, Petrillo M, Carone G, Sansone M, Franco R, Fulciniti F, Perdonà S. Multiparametric MRI for prostate cancer detection: Performance in patients with prostate-specific antigen values between 2.5 and 10 ng/mL. *J Magn Reson Imaging*. 2014;39:1206–12.
168. Ueno Y, Takahashi S, Kitajima K, Kimura T, Aoki I, Kawakami F, Miyake H, Ohno Y, Sugimura K. Computed diffusion-weighted imaging using 3-T magnetic resonance imaging for prostate cancer diagnosis. *Eur Radiol*. 2013;23(12):3509–16.
169. Morgan VA, Kyriazi S, Ashley SE, DeSouza NM. Evaluation of the potential of diffusion-weighted imaging in prostate cancer detection. *Acta radiol*. 2007;48(13):695–703.
170. Agha M, Eid AF. 3Tesla MRI surface coil: Is it sensitive for prostatic imaging?? *Alexandria J Med*. 2015;51(2):111–9.
171. Bains LJ, Studer UE, Froehlich JM, Giannarini G, Triantafyllou M, Fleischmann A, Thoeny HC. Diffusion-Weighted Magnetic Resonance Imaging Detects Significant Prostate Cancer with

- High Probability. *J Urol*. 2014;192:737–42.
172. Baur ADJ, Daqqaq T, Wagner M, Maxeiner A, Huppertz A, Renz D, Hamm B, Fischer T, Durmus T. T2- and diffusion-weighted magnetic resonance imaging at 3T for the detection of prostate cancer with and without endorectal coil: An intraindividual comparison of image quality and diagnostic performance. *Eur J Radiol*. 2016;85(6):1075–84.
173. Costa DN, Yuan Q, Xi Y, Rofsky NM, Lenkinski RE, Lotan Y, Roehrborn CG, Francis F, Travalini D, Pedrosa I. Comparison of prostate cancer detection at 3-T MRI with and without an endorectal coil: A prospective, Paired-patient study. *Urol Oncol Semin Orig Investig*. 2016;34(6):255.e7-255.e13.
174. Iwazawa J, Mitani T, Sassa S, Ohue S. Prostate cancer detection with MRI: Is dynamic contrast-enhanced imaging necessary in addition to diffusion-weighted imaging? *Diagnostic Interv Radiol*. 2011 Sep;17(3):243–8.
175. Kuhl CK, Bruhn R, Krämer N, Nebelung S, Heidenreich A, Schrading S. Abbreviated Biparametric Prostate MR Imaging in Men with Elevated Prostate-specific Antigen. *Radiology*. 2017;285(2):493–505.
176. Loggitsi D, Gyftopoulos A, Economopoulos N, Apostolaki A, Kalogeropoulos T, Thanos A, Alexopoulou E, Kelekis NL. Multiparametric Magnetic Resonance Imaging of the Prostate for Tumour Detection and Local Staging: Imaging in 1.5T and Histopathologic Correlation. *Can Assoc Radiol J*. 2017;68(4):379–86.
177. Rosenkrantz AB, Mannelli L, Kong X, Niver BE, Berkman DS, Babb JS, Melamed J, Taneja SS. Prostate cancer: Utility of fusion of T2-weighted and high b-value diffusion-weighted images for peripheral zone tumor detection and localization. *J Magn Reson Imaging*. 2011;34:95–100.
178. Rosenkrantz AB, Kim S, Campbell N, Gaing B, Deng F-M, Taneja SS. Transition Zone Prostate Cancer: Revisiting the Role of Multiparametric MRI at 3 T. *Am J Roentgenol*. 2015;204(3):W266–72.
179. Shimofusa R, Fujimoto H, Akamata H, Motoori K, Yamamoto S, Ueda T, Ito H. Diffusion-Weighted Imaging of Prostate Cancer. *J Comput Assist Tomogr*. 2005;29(2):149–53.
180. Shinmoto H, Tamura C, Soga S, Okamura T, Horiguchi A, Asano T, Kaji T. Anterior prostate

- cancer: Diagnostic performance of T2-weighted MRI and an apparent diffusion coefficient map. *Am J Roentgenol.* 2015 Jul 23;205(2):W185–92.
181. Stanzione A, Imbriaco M, Coccozza S, Fusco F, Rusconi G, Nappi C, Mirone V, Mangiapia F, Brunetti A, Ragozzino A, Longo N. Biparametric 3T Magentic Resonance Imaging for prostatic cancer detection in a biopsy-naïve patient population: a further improvement of PI-RADS v2? *Eur J Radiol.* 2016;85(12):2269–74.
182. Thestrup KCD, Logager V, Baslev I, Møller JM, Hansen RH, Thomsen HS. Biparametric versus multiparametric MRI in the diagnosis of prostate cancer. *Acta Radiol Open.* 2016;5(8):205846011666304.
183. Yoshimitsu K, Kiyoshima K, Irie H, Tajima T, Asayama Y, Hirakawa M, Ishigami K, Naito S, Honda H. Usefulness of apparent diffusion coefficient map in diagnosing prostate carcinoma: Correlation with stepwise histopathology. *J Magn Reson Imaging.* 2008;27(1):132–9.
184. Yoshizako T, Wada a, Hayashi T, Uchida K, Sumura M, Uchida N, Kitagaki H, Igawa M. Usefulness of diffusion-weighted imaging and dynamic contrast-enhanced magnetic resonance imaging in the diagnosis of prostate transition-zone cancer. *Acta radiol.* 2008;49(10):1207–13.
185. Tan CH, Wei W, Johnson V, Kundra V. Diffusion-weighted MRI in the detection of prostate cancer: meta-analysis. *AJR Am J Roentgenol.* 2012 Oct;199(4):822–9.
186. Vural M, Ertaş G, Onay A, Acar Ö, Esen T, Sağlıcan Y, Zengingönül HP, Akpek S. Conspicuity of peripheral zone prostate cancer on computed diffusion-weighted imaging: comparison of cDWI1500, cDWI2000, and cDWI3000. *Biomed Res Int.* 2014 Jan;2014:768291.
187. Ueno Y, Takahashi S, Ohno Y, Kitajima K, Yui M, Kassai Y, Kawakami F, Miyake H, Sugimura K. Computed Diffusion-Weighted MR Imaging for Prostate Cancer Detection: The Influence of the Combinations of b-Values. *Br J Radiol.* 2015;88(1048):20140738.
188. Augustin H, Erbersdobler A, Hammerer PG, Graefen M, Huland H. Prostate cancers in the transition zone: Part 2; clinical aspects. *BJU Int.* 2004;94(9):1226–9.
189. Greene DR, Wheeler TM, Egawa S, Weaver RP, Scardino PT. Relationship between clinical stage and histological zone of origin in early prostate cancer: morphometric analysis. *Br J*

- Urol. 1991 Nov;68(5):499–509.
190. Serefoglu EC, Altinova S, Ugras NS, Akincioglu E, Asil E, Balbay MD. How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer? *Can Urol Assoc J.* 2013;7(5–6):E293-8.
 191. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005;58(9):882–93.
 192. Rosenkrantz AB, Lim RP, Haghghi M, Somberg MB, Babb JS, Taneja SS. Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate MRI. *Am J Roentgenol.* 2013;201(4):W612–8.
 193. Muller BG, Shih JH, Sankineni S, Marko J, Rais-Bahrami S, George AK, De La Rosette JJMCH, Merino MJ, Wood BJ, Pinto P, Choyke PL, Turkbey B. Prostate cancer: Interobserver agreement and accuracy with the revised prostate imaging reporting and data system at multiparametric mr imaging1. *Radiology.* 2015;277(3):741–50.
 194. RANDOM.ORG - List Randomizer [Internet]. [cited 2017 Jun 6]. Available from: <https://www.random.org/lists/>
 195. Kitajima K, Kaji Y, Kuroda K, Sugimura K. High b-value diffusion-weighted imaging in normal and malignant peripheral zone tissue of the prostate: effect of signal-to-noise ratio. *Magn Reson Med Sci.* 2008 Jan;7(2):93–9.
 196. Park SY, Oh YT, Jung DC, Cho NH, Choi YD, Rha KH, Hong SJ. Diffusion-weighted imaging predicts upgrading of Gleason score in biopsy-proven low grade prostate cancers. *BJU Int.* 2017;119(1):57–66.
 197. Eng J. Sample Size Estimation: How Many Individuals Should Be Studied? *Radiology.* 2003;227(2):309–13.
 198. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12(1):77.
 199. Human Tissue Act 2004. <http://www.legislation.gov.uk/ukpga/2004/30/contents>; 2004.

200. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977 Mar;33(1):159.
201. Cicchetti D V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994 Jan 1;
202. Park SY, Kim CK, Park JJ, Park BK. Exponential apparent diffusion coefficient in evaluating prostate cancer at 3T: Preliminary experience. *Br J Radiol*. 2015;89(1058):no pagination.
203. Barrett T, Priest AN, Lawrence EM, Goldman DA, Warren AY, Gnanapragasam VJ, Sala E, Gallagher FA. Ratio of Tumor to Normal Prostate Tissue Apparent Diffusion Coefficient as a Method for Quantifying DWI of the Prostate. *AJR Am J Roentgenol*. 2015;205(6):W585-93.
204. Doo KW, Sung DJ, Park BJ, Kim MJ, Cho SB, Oh YW, Ko YH, Yang KS. Detectability of low and intermediate or high risk prostate cancer with combined T2-weighted and diffusion-weighted MRI. *Eur Radiol*. 2012;22(8):1812–9.
205. Rosenkrantz AB, Ginocchio LA, Cornfeld D, Froemming AT, Gupta RT, Turkbey B, Westphalen AC, Babb JS, Margolis DJ. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. *Radiology*. 2016 Sep;280(3):793–804.
206. Rosenkrantz AB, Chandarana H, Pfeuffer J, Triolo MJ, Shaikh MB, Mossa DJ, Geppert C. Zoomed echo-planar imaging using parallel transmission: impact on image quality of diffusion-weighted imaging of the prostate at 3T. *Abdom Imaging*. 2015;40(1):120–6.
207. Dale BM, Braithwaite AC, Boll DT, Merkle EM. Field strength and diffusion encoding technique affect the apparent diffusion coefficient measurements in diffusion-weighted imaging of the abdomen. *Invest Radiol*. 2010;45(2):104–8.
208. Bülow R, Mensel B, Meffert P, Hernando D, Evert M, Kühn JP. Diffusion-weighted magnetic resonance imaging for staging liver fibrosis is less reliable in the presence of fat and iron. *Eur Radiol*. 2013;23(5):1281–7.
209. Isebaert S, Van Den Bergh L, Haustermans K, Lerut E, De Wever L, Budiharto T, Slagmolen P, Joniau S, Van Poppel H, Claus F, Oyen R. Multi-modality MR imaging for prostate cancer detection in correlation to whole-mount histopathology. *Radiother Oncol*. 2011;99:S129.
210. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research.

- Lancet (London, England). 1991 Apr 13;337(8746):867–72.
211. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess*. 2000;4(10):1–115.
212. Osugi K, Tanimoto A, Nakashima J, Shinoda K, Hashiguchi A, Oya M, Jinzaki M, Kuribayashi S. What is the most effective tool for detecting prostate cancer using a standard MR scanner? *Magn Reson Med Sci*. 2013;12(4):271–80.
213. Turkbey B, Merino MJ, Gallardo EC, Shah V, Aras O, Bernardo M, Mena E, Daar D, Rastinehad AR, Linehan WM, Wood BJ, Pinto P a., Choyke PL. Comparison of endorectal coil and nonendorectal coil T2W and diffusion-weighted MRI at 3 Tesla for localizing prostate cancer: Correlation with whole-mount histopathology. *J Magn Reson Imaging*. 2014;39(6):1443–8.
214. Ahmed HU. The Index Lesion and the Origin of Prostate Cancer. *N Engl J Med*. 2009;361(17):1704–6.
215. Bott SRJ, Ahmed HU, Hindley RG, Abdul-Rahman A, Freeman A, Emberton M. The index lesion and focal therapy: An analysis of the pathological characteristics of prostate cancer. *BJU Int*. 2010;106(11):1607–11.
216. Mazaheri Y, Afaq AA, Jung SI, Goldman DA, Wang L, Aslan H, Zelefsky MJ, Akin O, Hricak H. Volume and landmark analysis: Comparison of MRI measurements obtained with an endorectal coil and with a phased-array coil. *Clin Radiol*. 2015;70(4):379–86.
217. desouza NM, Reinsberg SA, Scurr ED, Brewster JM, Payne GS. Magnetic resonance imaging in prostate cancer: the value of apparent diffusion coefficients for identifying malignant nodules. *Br J Radiol*. 2007 Feb;80(950):90–5.
218. Rosenkrantz AB, Triolo MJ, Melamed J, Rusinek H, Taneja SS, Deng FM. Whole-lesion apparent diffusion coefficient metrics as a marker of percentage Gleason 4 component within gleason 7 prostate cancer at radical prostatectomy. *J Magn Reson Imaging*. 2015;41(3):708–14.
219. Donati OF, Afaq A, Vargas HA, Mazaheri Y, Zheng J, Moskowitz CS, Hricak H, Akin O. Prostate MRI: Evaluating tumor volume and apparent diffusion coefficient as surrogate biomarkers for predicting tumor Gleason score. *Clin Cancer Res*. 2014;20(14):3705–11.

220. Kim TH, Jeong JY, Lee SW, Kim CK, Park BK, Sung HH, Jeon HG, Jeong BC, Seo S II, Lee HM, Choi HY, Jeon SS. Diffusion-weighted magnetic resonance imaging for prediction of insignificant prostate cancer in potential candidates for active surveillance. *Eur Radiol*. 2015 Jun;25(6):1786–92.
221. Kim TH, Kim TH, Kim CK, Park BK, Jeon HG, Jeong BC, Seo S II, Lee HM, Choi HY, Jeon SS. Relationship between Gleason score and apparent diffusion coefficients of diffusion-weighted magnetic resonance imaging in prostate cancer patients. *Can Urol Assoc J*. 2016;10(11–12):E377-82.
222. Wang L, Mazaheri Y, Zhang J, Ishill NM, Kuroiwa K, Hricak H. Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy. *Radiology*. 2008;246(1):168–76.
223. Qayyum A, Coakley F V., Lu Y, Olpin JD, Wu L, Yeh BM, Carroll PR, Kurhanewicz J. Organ-Confined Prostate Cancer : Effect of Prior Transrectal Biopsy on Endorectal MRI and Spectroscopic imaging. *Am J Roentgenol*. 2004;183(4):1079–83.
224. Tamada T, Sone T, Jo Y, Yamamoto A, Yamashita T, Egashira N, Imai S, Fukunaga M. Prostate cancer: relationships between postbiopsy hemorrhage and tumor detectability at MR diagnosis. *Radiology*. 2008 Aug;248(2):531–9.
225. Park KK, Lee SH, Lim BJ, Kim JH, Chung BH. The effects of the period between biopsy and diffusion-weighted magnetic resonance imaging on cancer staging in localized prostate cancer. *BJU Int*. 2010;106(8):1148–51.
226. Ikonen S, Kivisaari L, Vehmas T, Tervahartiala P, Salo JO, Taari K, Rannikko S. Optimal timing of post-biopsy MR imaging of the prostate. *Acta Radiol*. 2001;42(1):70–3.

Appendix 1

Published material and presentations

Godley KC, Syer TJ, Toms AP, Smith TO, Johnson G, Cameron D, Malcolm PN. Accuracy of high b-value diffusion-weighted MRI for prostate cancer detection: A meta-analysis. *Acta Radiol.* 2017;doi: 10.1177/0284185117702181

Original Article

Accuracy of high b-value diffusion-weighted MRI for prostate cancer detection: a meta-analysis

Keith Craig Godley¹, Tom Joseph Syer², Andoni Paul Toms¹, Toby Oliver Smith², Glyn Johnson², Donnie Cameron² and Paul Napier Malcolm¹

Acta Radiologica
0(0) 1–9
© The Foundation Acta Radiologica
2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0284185117702181
journals.sagepub.com/home/acr
SAGE

Abstract

Background: The diagnostic accuracy of diffusion-weighted imaging (DWI) to detect prostate cancer is well-established. DWI provides visual as well as quantitative means of detecting tumor, the apparent diffusion coefficient (ADC). Recently higher b-values have been used to improve DWI's diagnostic performance.

Purpose: To determine the diagnostic performance of high b-value DWI at detecting prostate cancer and whether quantifying ADC improves accuracy.

Material and Methods: A comprehensive literature search of published and unpublished databases was performed. Eligible studies had histopathologically proven prostate cancer, DWI sequences using b-values ≥ 1000 s/mm², less than ten patients, and data for creating a 2 × 2 table. Study quality was assessed with QUADAS-2 (Quality Assessment of diagnostic Accuracy Studies). Sensitivity and specificity were calculated and tests for statistical heterogeneity and threshold effect performed. Results were plotted on a summary receiver operating characteristic curve (sROC) and the area under the curve (AUC) determined the diagnostic performance of high b-value DWI.

Results: Ten studies met eligibility criteria with 13 subsets of data available for analysis, including 522 patients. Pooled sensitivity and specificity were 0.59 (95% confidence interval [CI], 0.57–0.61) and 0.92 (95% CI, 0.91–0.92), respectively, and the sROC AUC was 0.92. Subgroup analysis showed a statistically significant ($P=0.03$) improvement in accuracy when using tumor visual assessment rather than ADC.

Conclusion: High b-value DWI gives good diagnostic performance for prostate cancer detection and visual assessment of tumor diffusion is significantly more accurate than ROI measurements of ADC.

Keywords

Prostate cancer, magnetic resonance imaging, MRI, diffusion-weighted imaging, DWI, high-b-value, meta-analysis

Date received: 2 August 2016; accepted: 25 February 2017

Introduction

Prostate cancer is the most commonly diagnosed cancer and the second most common cause of cancer-related death in men (1,2). Magnetic resonance imaging (MRI) is the imaging mainstay of prostate cancer localization, being recommended in men considered for radical treatment following positive trans-rectal ultrasound (TRUS) biopsy, and in high-risk patients with a negative biopsy, active surveillance, response to treatment, and recurrence (3,4).

Multi-parametric MRI of the prostate comprises T1-weighted (T1W) and T2-weighted (T2W) imaging,

with additional techniques such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced MRI (3). DWI has become a routine part of prostate MRI protocols, as it provides good tumor contrast without exogenous agents.

¹Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK

²University of East Anglia, Norwich, UK

Corresponding author:

Keith Craig Godley, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Lane, Norwich NR4 7UY, UK.
Email: kcgodley@googlemail.com

Applying DWI with multiple diffusion weightings, or b-values, allows the apparent diffusion coefficient (ADC) to be estimated: a parameter known to inversely correlate with tumor aggressiveness (5,6). Clinically, both DW images, typically higher b-values, and ADC maps are assessed to detect tumor, which appears bright on diffusion images and dark on ADC maps.

ADC maps are calculated through mono-exponential fits to diffusion signal decays on a voxel-by-voxel basis. Other signal models, have been applied to prostate DWI, including diffusional kurtosis (7), intravoxel incoherent motion (8) the stretched exponential (9), and VERDICT (10). However, the mono-exponential model is the most common, requiring only two b-values for fitting.

Performing mono-exponential fitting with higher maximum b-values, b_{max} , can improve contrast-to-noise (C/N) in the resulting maps, where C/N is defined as: $(\text{Signal}(\text{lesion}) - \text{Signal}(\text{background})) / \text{estimated noise}$. This gives better characterization of ADC differences between normal and cancerous prostate, improving tumor detection at the expense of reduced signal-to-noise (S/N) and increased sensitivity to motion artifacts. At 1.5T, b_{max} of 500 and 1000 s/mm² are typically used, but increased S/N at higher field strengths permits the use of higher b_{max} (11). Guidelines recommend a b_{max} of 800–1000 s/mm² (3); but there is no consensus on whether higher b-values ($b_{max} \geq 1000$ s/mm²) should be used routinely. Many studies have compared high and lower b-value ($b_{max} < 1000$ s/mm²) acquisitions, but results have been conflicting (12–17).

The diagnostic accuracy of DWI is well-established, with multiple meta-analyses reporting its diagnostic performance (18–20). There is uncertainty about the usefulness of high b-value DWI, particularly which b-value provides ADC maps with the greatest diagnostic performance. The aim of this study is to determine the diagnostic performance of high b-value DWI at detecting prostate cancer and whether quantifying ADC improves accuracy.

Material and Methods

This meta-analysis was reported using the preferred reporting items for systematic reviews and meta-analyses outlined in the PRISMA statement (21). The review was registered prior to commencing on PROSPERO (ref no. CRD42015027644).

Search strategy

A comprehensive systematic literature search was independently performed by two reviewers (KCG, TSy) to identify studies investigating the diagnostic accuracy or

performance of high b-value DWI for detecting prostate cancer. A MEDLINE search is presented in Supplementary Table 1. In addition, searches were conducted of EMBASE, and the gray literature/trial registry databases: WHO International Clinical Trials Registry Platform and OpenGrey. Studies were not limited by country of origin, but were limited to those published in English. All searches were from database inception to 1 January 2016.

Eligibility criteria

Retrospective and prospective studies were included if they reported detection of prostate cancer in a pre-treatment population using high b-value DWI of the prostate. Only primary research articles, available as full-text, were accepted; however, review articles were checked for additional primary references. High b-value was defined as $b_{max} \geq 1000$ s/mm². Histopathological results as a reference standard (biopsy or radical prostatectomy), sufficient data to calculate true positive (TP), false positive (FP), false negative (FN), and true negative (TN) data were required. If multiple b-values were used, including $b < 1000$ s/mm², the study was only eligible if data with $b \geq 1000$ s/mm² could be extracted. Studies using high b-value DWI in combination with other diagnostic sequences to detect cancer were excluded.

Study identification

Titles and abstracts from the search results, and the full-text papers for all studies which met or potentially met the eligibility criteria, were independently reviewed by two reviewers (KCG, TSy). Those studies which met the eligibility criteria on full review were included in the final analysis. Disagreements on inclusion suitability were resolved by consensus (KCG, TSy).

Data extraction

Two reviewers (KCG, TSy) independently extracted the data on a pre-defined template, including: publication year, country of origin, sample size, description of study population (age), study design (prospective, retrospective, or unknown), patient enrollment (consecutive or not), inclusion and exclusion criteria, reasons for exclusions from analysis, and number of experts who assessed and interpreted MRI results. Data were recorded on: blinding of MRI measurements to clinical, biochemical, or histopathological results; methods used to determine diagnosis; types of coils; and b-values used. For each study, we recorded the number of TP, FP, TN, and FN findings for high b-value DWI in diagnosing prostate cancer.

Disagreements in data extraction findings were resolved through discussion or through adjudication with a third reviewer (TSM).

Quality assessment

Two reviewers (KCG, TSy) independently assessed each included paper's quality using QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) (22). Any disagreements were resolved through discussion or through adjudication with a third reviewer (TSM).

Statistical analysis

Study heterogeneity was assessed through examination of the data extract table. This indicated broad study homogeneity, meaning a meta-analysis was appropriate. Statistical heterogeneity was assessed using the chi-squared statistic, I^2 , and the inconsistency, I^2 . When $P < 0.10$ and $I^2 > 50\%$, unexplained statistical heterogeneity was evident and diagnostic performance analyses were performed using a random-effects model.

Specificity and sensitivity of each study was calculated using 2×2 contingency tables. Pooled sensitivity and specificity and positive and negative likelihood ratios with 95% confidence intervals (CIs) were calculated. Finally, the specificity and sensitivity were used to calculate a summary receiver operating characteristic (sROC) curve and the area under the curve (AUC).

The threshold effect was assessed visually, by determining whether the ROC curve presented with a "shoulder-arm" shape, and qualitatively using the Spearman correlation coefficient of the logit of sensitivity and the logit of (1-specificity), with $P < 0.05$ indicating the heterogeneity between studies could not be explained by threshold effect. A meta-regression and subgroup analysis was performed to explore other

sources of heterogeneity and how they influence diagnostic performance.

All statistical computations were performed using Meta-DiSc (version 1.4, Javier Zamora) and Review Manager (version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014).

Results

Search results

A summary of the search strategy results is presented in Supplementary Fig. 1. In total, 351 studies were identified from the search results, of which 61 were deemed potentially eligible. After full-text review, ten studies met the final eligibility criteria and were included in the analysis (16,17,23–30).

Characteristics of included studies

The principle characteristics of the included studies are displayed in Table 1, with imaging and study methods listed in Table 2. From the ten included studies, 522 patients were analyzed, with a mean age of 64 years (age range, 43–87 years). The mean and median prostate-specific antigens were 19 and 9.3 ng/mL, respectively.

Three studies were prospective and seven retrospective. Field strengths of 1.5T (23,25,26,28,29) and 3T (16,17,24,27,30) were each used in five studies. Radical prostatectomy specimens were used as the reference standard in six studies (16,17,25,27,29,30), biopsy specimens in three (23,24,26), and one study used a combination (28). The MRI reader was blinded in eight studies (16,17,23,25–28,30); blinding was not known in two. Anti-spasmodic agents, either glucagon or hyoscine butylbromide, were used in five studies (16,17,24,29,30) and their use unknown in the

Table 1. Principle characteristics of included studies.

Study	Year	Country	Patients (n)	Age (years) (range)	PSA (range)	Design
Chen	2008	China	42	63 (45–82)	52.5 (4.7–147)	Retro
Girometti	2012	Italy	26	64* (51–74)	6.0* (2.5–10)	Pro
Isebaert	2013	Belgium	75	66* (49–74)	10.4 (1.5–70.9)	Pro
Kim	2000	Republic of Korea	48	66 (45–80)	7.2* (2.3–23.2)	Retro
Koo	2013	Republic of Korea	80	66 (45–81)	7.2 (1.2–57)	Retro
Kumar	2007	India	23	64.5	11* (0.5–1000)	Pro
Lim	2009	Republic of Korea	52	65 (48–76)	66 (45–80)	Retro
Peng	2013	USA	48	61.5 (44–73)	15.6 (0.8–256)	Retro
Rosenkrantz	2015	USA	58	63	8.2	Retro
Vilanova	2011	Spain	70	63.5 (43–87)	7.4* (4–17.2)	Retro

*Median.

PSA, prostate specific antigen (ng/mL); Pro, prospective; Retro, retrospective.

Table 2. Imaging and methodological characteristics of the included studies.

Study	Field strength	Coil*	b-values (s/mm ²)	TR/TE (ms)	NSA	Voxel size (mm)	RS	Blind	Spasm	Threshold used
Chen	1.5T	A	0, 1000	3200/94	–	1.8 × 1.8 × –	Bx	Y	U	ADC
Girometti	3T	A	0, 800, 1200	4600/60	6	2.3 × 2.5 × 3.0	Bx	U	Y	ADC
Isebaert	1.5T	A	0, 50, 100, 500, 750, 1000	4000/79	–	3.0 × 3.0 × 4.0	RP	Y	U	Visual
Kim	3T	A	0, 1000; 0, 2000	2740–2950/83–95	3	1.8 × 1.8 × 3.0	RP	Y	Y	Visual
Koo	3T	A	0, 300; 0, 700; 0, 1000; 0, 2000	4830–4840/75–76	3	1.8 × 1.8 × 3.0	RP	Y	Y	Visual scale
Kumar	1.5T	B	0, 250, 500, 750, 1000	3000/96	–	1.8 × 1.8 × 4.0–5.0	Bx	Y	U	ADC
Lim	3T	A	0, 1000	3400/117	–	0.9 × 0.9 × 4.0	RP	Y	Y	Visual scale
Peng	1.5T	B	0, 50, 200, 1500, 2000; 0, 1000 [†]	2948–8616/71–85	1–4	0.81 × 0.81–1.28 × 1.28	RP	U	Y	ADC
Rosenkrantz	3T	A	50, 1000	4100/86	10	2.0 × 0.9 × 3.0	RP	Y	U	ADC
Vilanova	1.5T	B	0, 1000	8250/94	6	2.0 × 2.0 × 3.0	C	Y	U	ADC

*Twenty-nine patients were imaged with b-values of 0, 50, 200, 1500, and 2000; 24 patients were imaged with b-values of 0 and 1000.

[†]Coil A = without endorectal coil, coil B = with endorectal coil.

ADC, apparent diffusion coefficient; Blind, blinded; Bx, biopsy; C, both RP and Bx included; NSA, number of signal averages; RP, radical prostatectomy; RS, reference standard; Spasm, anti-spasmodics; TE, echo time; Th, threshold; TR, repetition time; U, unclear; Y, yes.

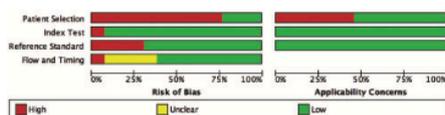


Fig. 1. QUADAS-2 results summarizing the proportion of low, high, and unclear risk of bias and applicability concerns.

remainder. Five studies (16,17,24,29,30) used b-values > 1000 s/mm².

Several methods were used to detect prostate cancer: region-of-interest (ROI)-based ADC quantification was used in six studies (23,24,26–29) and visual assessment of lesions using ADC maps was performed in four studies (16,17,25,30), of which two (17,30) used a scale (such as Likert scale), and the other two used a binary cutoff.

All ten studies used the mono-exponential function only to estimate ADC. In three studies (16,17,27), extracted data were split into subsets. Kim et al. (16) and Koo et al. (17) generated multiple ADC maps: b = (0, 1000), and (0, 2000) s/mm² for the former, and b = (0, 300), (0, 700), (0, 1000), and (0, 2000) s/mm² for the latter. Rosenkrantz et al. (27) split results from the peripheral zone and the transitional zone. The other studies generated only one set of ADC maps from their DWI data, performing mono-exponential fitting to all acquired b-values.

Quality assessment

Study quality assessment is presented in Supplementary Table 2. Fig. 1 demonstrates the QUADAS-2 graphical summary of the studies indicating the proportion of

high, low, or unclear risk in each domain. A high risk of bias was demonstrated in the patient selection domain, but overall the quality of the studies included was considered “good.”

Diagnostic performance

Diagnostic results of the meta-analyses are presented in Table 3. The results from the individual studies are presented in Supplementary Table 3.

The pooled sensitivity and specificity of high b-value DWI MRI in detecting prostate cancer were 0.59 (95% CI, 0.57–0.61; Fig. 2) and 0.92 (95% CI, 0.91–0.92; Fig. 3), respectively. Sensitivity and specificity heterogeneity tests gave $Q = 435.05$ ($P < 0.001$), $I^2 = 97.2\%$ and $Q = 89$ ($P < 0.001$), $I^2 = 86.5\%$ respectively, indicating significant statistical heterogeneity between studies.

The pooled positive and negative likelihood ratios for high b-value DWI MRI in detecting prostate cancer were 6.64 (95% CI, 4.9–9.0; Supplementary Fig. 2) and 0.33 (95% CI, 0.2–0.5; Supplementary Fig. 3), respectively. Positive and negative likelihood ratio heterogeneity tests gave $Q = 82.50$ ($P < 0.001$), $I^2 = 85.5\%$ and $Q = 517.45$ ($P < 0.001$), $I^2 = 97.7\%$, respectively, indicating significant statistical heterogeneity between studies.

Fig. 4 shows the sROC curve of the ten studies, where AUC = 0.92, indicating “good” diagnostic accuracy (31).

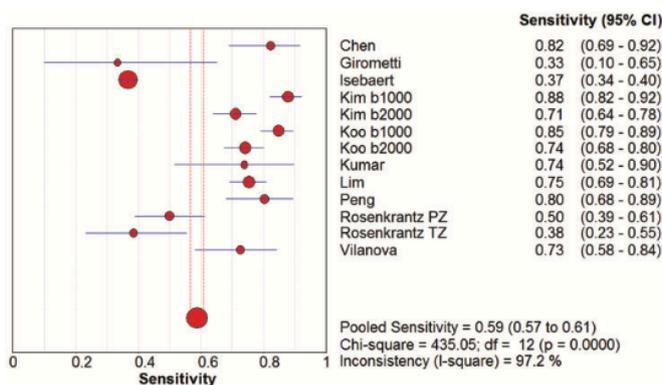
Meta-regression analysis

The ROC curve did not demonstrate a “shoulder-arm” shape (Supplementary Table 4) and the Spearman

Table 3. Results of the subgroup analysis.

Study characteristics	n	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	AUC	P value*
Total	13	0.59 (0.57–0.91)	0.92 (0.91–0.92)	0.92	
b-value					0.31
1000	9	0.55 (0.53–0.58)	0.90 (0.89–0.92)	0.91	
> 1000	3	0.72 (0.67–0.76)	0.94 (0.93–0.95)	0.98	
Field strength					0.20
1.5T	6	0.49 (0.46–0.51)	0.90 (0.88–0.91)	0.89	
3T	7	0.74 (0.71–0.79)	0.93 (0.92–0.94)	0.96	
Coil					0.32
With endorectal	4	0.76 (0.71–0.80)	0.86 (0.83–0.89)	0.84	
Without endorectal	9	0.56 (0.53–0.58)	0.93 (0.92–0.93)	0.94	
Reference standard					0.16
Biopsy	3	0.73 (0.63–0.82)	0.86 (0.82–0.89)	0.86	
Prostatectomy	9	0.56 (0.56–0.60)	0.92 (0.91–0.93)	0.94	
Threshold method					0.03
ADC	7	0.64 (0.59–0.69)	0.87 (0.84–0.89)	0.88	
Visual	6	0.58 (0.55–0.60)	0.93 (0.92–0.93)	0.95	
Patient selection bias					0.29
High risk	10	0.58 (0.56–0.60)	0.92 (0.91–0.93)	0.94	
Low risk	3	0.77 (0.68–0.84)	0.83 (0.79–0.87)	0.86	

*P value = comparison of diagnostic odds ratio of subgroups.
AUC, area under the curve.

**Fig. 2.** Forest plot of sensitivity with pooled sensitivity, Q statistic of the chi-squared, and I-squared results.

Correlation Coefficient between the logit of sensitivity and the logit of (1-specificity) was 0.286 ($P=0.344$), confirming that the threshold effect is not responsible for the variation in accuracy between studies.

Subgroup analysis

Subgroup analysis was based on different study characteristics and perceived sources of bias and

applicability uncovered in the QUADAS assessment. Studies at 3T with and without an endorectal coil demonstrated the highest pooled sensitivity of 0.76 (95% CI, 0.71–0.80) and 0.74 (95% CI, 0.71–0.79) respectively. When assessing protocols with a $b_{\max} > 1000 \text{ s/mm}^2$, the pooled specificity and AUC of the sROC were greater: 0.94 (95% CI, 0.93–0.95) and 0.98, respectively. A statistically significant ($P < 0.05$) improvement was seen using assessment of tumor

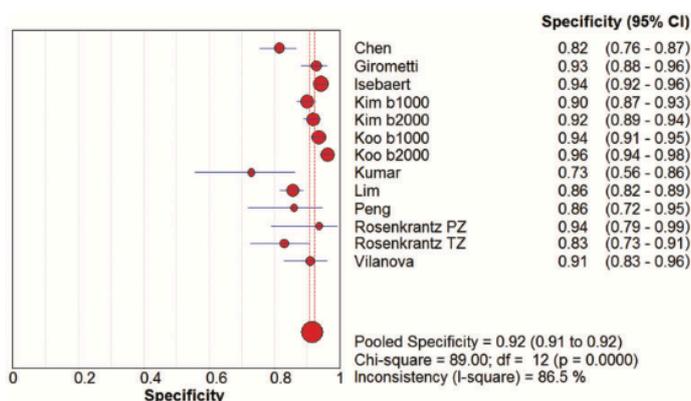


Fig. 3. Forest plot of specificity with pooled specificity, Q-statistic of the chi-squared, and I-squared results.

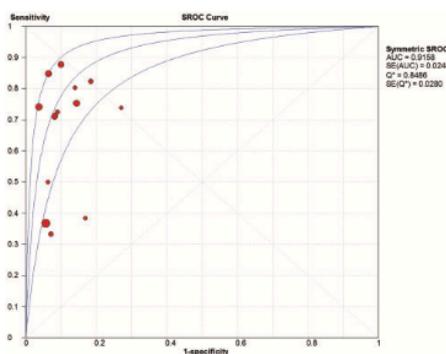


Fig. 4. The sROC curve for high b-value DWI in detecting prostate cancer.

presence on ADC maps as a visual threshold versus ROI measurements. The diagnostic performance of the subgroup analysis and *P* values of the above-mentioned factors and others are demonstrated in Table 3.

Discussion

This analysis indicates that high b-value imaging is a good diagnostic tool for detecting prostate cancer. The results of the threshold method subgroup analysis imply that there is a benefit in using higher b_{\max} in a clinical setting. The lesser value of quantitative ADC thresholding as a tool for detecting tumor is in line with PI-RADS version 2 recommendations (standardized reporting standards for prostate MRI) (32). The evidence on which this analysis was made was graded as “good” quality using the QUADAS-2 tool.

There have been multiple meta-analyses investigating the diagnostic accuracy of DWI alone or in

combination with other imaging techniques (18–20). The pooled sensitivity, specificity, and overall accuracy of our study were 0.58, 0.92, and 0.92, respectively, similar to Jie et al.’s (19) meta-analysis of DWI alone. This is likely due to overlap of included studies, with nine of the ten included studies featured in their meta-analysis. However, in contrast to Jin et al.’s meta-analysis of all b-values, the sensitivity was lower in our study (0.58 versus 0.77), but the pooled specificity and AUC were higher (0.92 versus 0.84 and 0.92 versus 0.88 respectively) (20). This suggests high b-value imaging may help to rule out significant prostate cancer.

There was significant statistical heterogeneity between the included studies that could not be explained by threshold effect. Subgroup analyses of multiple study parameters were assessed to attempt to explain the heterogeneity. Given the unknown cause of statistical heterogeneity, these findings should be interpreted cautiously.

Improved tumor contrast at high b-values comes at the cost of decreased S/N (11); however, this can be mitigated through the use of 3T field strength. Most of the diagnostically specific high b-value diagnostic accuracy studies use 3T (12,33–35). The subgroup analysis of field strength demonstrated a trend towards improved accuracy with 3T. The sensitivity results of the 3T group alone are similar and the specificity and accuracy are better than those found in a meta-analysis of accuracy of visual assessment of combined T2 and DWI sequences for prostate cancer detection by Wu et al. (21).

On review of the method of prostate cancer detection, about half of the studies qualitatively assessed the ability of blinded readers to visually detect prostate cancer either by answering a binary question regarding cancer presence or using a probability scale (16,17,25,30). The remaining studies used ROI-based

ADC calculations to determine prostate cancer presence or absence, or used a scale of ADC values to predict cancer (23,24,26–29). The diagnostic performance of tumor visual assessment on ADC maps was significantly better than quantitative ADC methods, with visual assessment giving similar results to Wu et al.'s combined T2 and DWI meta-analysis (18), indicating potential value for high b-value imaging clinically.

A potential explanation for the relatively poor performance of quantitative ADC methods is four of the six studies included used biopsy as a reference standard whereas all visual assessment studies used radical prostatectomy specimens. TRUS biopsy as a reference standard is limited given its poor octant localization, small sampling volume, and substantial number of tumors missed (36,37). Radical prostatectomy studies performed better than biopsy studies, but the difference was not statistically significant.

ADC estimation is influenced by a number of factors, including, but not limited to: noise, fat, and perfusion signals (8,38,39); possible non-Gaussian diffusion (7); and the known diffusion anisotropy of the prostate (40). Because of these confounds, other diffusion models may ultimately prove more appropriate for identifying prostate cancer (41,42). Along with the study-specific TE, diffusion gradient duration, δ , and diffusion time, Δ , such factors could substantially influence the sensitivity and specificity of DWI for evaluating prostate cancer. None of the studies reviewed reported δ and Δ ; few researchers (42,43) have considered these factors when applying prostate DWI.

None of the other subgroup analyses demonstrated a significant difference between the groups. In some subgroups, a statistical difference would have been difficult to demonstrate given the small numbers of studies. For example, only three studies used biopsy as a reference standard, but despite this, this subgroup analysis provided the second strongest statistical source of heterogeneity ($P=0.18$). The limitations of biopsy as a reference standard are described above, but limiting MRI assessment to patients who have had a prostatectomy introduces patient selection bias, as radical prostatectomy patients tend to be younger, fitter, and tend to have clinically significant tumor, prompting surgery. Radical prostatectomy allows examination of the entire gland including the anterior gland (which TRUS cannot) and detects multifocality, which is frequent (44,45). Eight of nine studies in the prostatectomy subgroup assessed for multifocal disease or for tumor in multiple segments of the prostate and this subgroup's results may be more representative of the diagnostic accuracy of high b-value diffusion.

There are limitations to this study. Limitation by language and database may have introduced bias. The use of two larger databases, and grey literature, should

encompass most eligible English language studies. Publication bias was not assessed as a Deek's funnel plot is less accurate in meta-analyses with small numbers of studies (46). Finally, this study was restricted to testing localization of prostate cancer within the gland. This is important in determining the accuracy of high b-value diffusion, but not the only useful outcome. Identifying the presence of capsular breach, seminal vesicle invasion and pelvic lymphadenopathy are important staging and prognostic characteristics not assessed in this study.

In conclusion, these findings should be considered cautiously given the degree of statistical heterogeneity. However, this meta-analysis demonstrated that high b-value diffusion is a valuable diagnostic tool, with a sensitivity of 59%, specificity of 92% and sROC AUC of 0.92. There was better diagnostic performance by visual assessment of high b-value DWI studies compared to ADC quantification.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer Statistics. *CA Cancer J Clin* 2009;59:225–249.
2. American Cancer Society. *Cancer Facts & Figures 2012*. Atlanta, GA: American Cancer Society, 2012.
3. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. *Eur Radiol* 2012;22:746–757.
4. Kirkham A, Haslam P, Keanie JY, et al. Prostate MRI: who, when, and how? Report from a UK consensus meeting. *Clin Radiol* 2013;68:1016–1023.
5. Tamada T, Sone T, Jo Y, et al. Apparent diffusion coefficient values in peripheral and transition zones of the prostate: comparison between normal and malignant prostatic tissues and correlation with histologic grade. *J Magn Reson Imaging* 2008;28:720–726.
6. Hambrock T, Somford DM, Huisman HJ, et al. Relationship between apparent diffusion coefficients at 3.0-T mr imaging and gleason grade in peripheral zone prostate cancer. *Radiology* 2011;259:453–461.
7. Rosenkrantz AB, Sigmund EE, Johnson G, et al. Prostate cancer: Feasibility and preliminary experience of a diffusional kurtosis model for detection and assessment of aggressiveness of peripheral zone cancer. *Radiology* 2012;264:126–135.
8. Riches SF, Hawtin K, Charles-Edwards EM, et al. Diffusion-weighted imaging of the prostate and rectal wall: Comparison of biexponential and monoexponential

- modelled diffusion and associated perfusion coefficients. *NMR Biomed* 2009;22:318–325.
9. Mazaheri Y, Afaq A, Rowe DB, et al. Diffusion-weighted magnetic resonance imaging of the prostate: Improved robustness with stretched exponential modeling. *J Comput Assist Tomogr* 2012;36:695–703.
 10. Panagiotaki E, Chan RW, Dikaios N, et al. Microstructural characterization of normal and malignant human prostate tissue with vascular, extracellular, and restricted diffusion for cytometry in tumours magnetic resonance imaging. *Invest Radiol* 2015;50:218–227.
 11. Kim CK, Park BK, Kim B. Diffusion-weighted MRI at 3 T for the evaluation of prostate cancer. *Am J Roentgenol* 2010;194:1461–1469.
 12. Wang X, Qian Y, Liu B, et al. High-b-value diffusion-weighted MRI for the detection of prostate cancer at 3 T. *Clin Radiol* 2014;69:1165–1170.
 13. Ueno Y, Kitajima K, Sugimura K, et al. Ultra-high b-value diffusion-weighted MRI for the detection of prostate cancer with 3-T MRI. *J Magn Reson Imaging* 2013;38:154–160.
 14. Tamada T, Kanomata N, Sone T, Jo Y, et al. High b value (2,000 s/mm²) diffusion-weighted magnetic resonance imaging in prostate cancer at 3 Tesla: comparison with 1,000 s/mm² for tumor conspicuity and discrimination of aggressiveness. *PLoS One* 2014;9:e96619.
 15. Rosenkrantz AB, Hindman N, Lim RP, et al. Diffusion-weighted imaging of the prostate: Comparison of b1000 and b2000 image sets for index lesion detection. *J Magn Reson Imaging* 2013;38:694–700.
 16. 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 s/mm². *Am J Roentgenol* 2010;194:W33–37.
 17. Koo JH, Kim CK, Choi D, et al. Diffusion-weighted magnetic resonance imaging for the evaluation of prostate cancer: Optimal B value at 3T. *Korean J Radiol* 2013;14:61–69.
 18. Wu L-M, Xu J-R, Ye Y-Q, et al. The clinical value of diffusion-weighted imaging in combination with T2-weighted imaging in diagnosing prostate carcinoma: a systematic review and meta-analysis. *Am J Roentgenol* 2012;199:103–110.
 19. Jie C, Rongbo L, Ping T, et al. The value of diffusion-weighted imaging in the detection of prostate cancer: a meta-analysis. *Eur Radiol* 2014;24:1929–1941.
 20. Jin G, Su DK, Luo N Bin, et al. Meta-analysis of diffusion-weighted magnetic resonance imaging in detecting prostate cancer. *J Comput Assist Tomogr* 2013;37:195–202.
 21. Moher D, Shamseer L, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
 22. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–536.
 23. Chen M, Dang H-D, Wang J-Y, et al. Prostate cancer detection: comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined. *Acta Radiol* 2008;49:602–610.
 24. Girometti R, Bazzocchi M, Como G, et al. Negative predictive value for cancer in patients with “gray-Zone” PSA level and prior negative biopsy: Preliminary results with multiparametric 3.0 tesla MR. *J Magn Reson Imaging* 2012;36:943–950.
 25. Isebaert S, Van den Bergh L, Haustermans K, et al. Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology. *J Magn Reson Imaging* 2013;37:1392–1401.
 26. Kumar V, Jagannathan NR, Kumar R, et al. Apparent diffusion coefficient of the prostate in men prior to biopsy: determination of a cut-off value to predict malignancy of the peripheral zone. *NMR Biomed* 2007;20:505–511.
 27. Rosenkrantz ABB, Khalef V, Xu W, et al. Does normalisation improve the diagnostic performance of apparent diffusion coefficient values for prostate cancer assessment? A blinded independent-observer evaluation. *Clin Radiol* 2015;70:1032–1037.
 28. Vilanova JC, Barceló-Vidal C, Comet J, et al. Usefulness of prebiopsy multifunctional and morphologic MRI combined with free-to-total prostate-specific antigen ratio in the detection of prostate cancer. *Am J Roentgenol* 2011;196:715–722.
 29. Peng Y, Jiang Y, Yang C, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study. *Radiology* 2013;267:787–796.
 30. Lim HK, Kim JK, Kim K, et al. Prostate cancer: Apparent diffusion coefficient map with T2-weighted images for detection - A multireader study. *Radiology* 2009;250:145–151.
 31. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg* 2005;79:16–20.
 32. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging-Reporting and Data System: 2015, Version 2. *Eur Urol* 2016;69:16–40.
 33. Ueno Y, Kitajima K, Sugimura K, et al. Ultra-high b-value diffusion-weighted MRI for the detection of prostate cancer with 3-T MRI. *J Magn Reson Imaging* 2013;38:154–160.
 34. Kitajima K, Takahashi S, Ueno Y, et al. Do apparent diffusion coefficient (ADC) values obtained using high b-values with a 3-T MRI correlate better than a transrectal ultrasound (TRUS)-guided biopsy with true Gleason scores obtained from radical prostatectomy specimens for patients with prostat. *Eur J Radiol* 2013;82:1219–1226.
 35. Ohgiya Y, Suyama J, Seino N, et al. Diagnostic accuracy of ultra-high-b-value 3.0-T diffusion-weighted MR imaging for detection of prostate cancer. *Clin Imaging* 2012;36:526–531.
 36. Catalona WJ, Partin AW, Slawin KM, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic

- disease: a prospective multicenter clinical trial. *JAMA* 1998;279:1542–1547.
37. Ahmed HU, Kirkham A, Arya M, et al. Is it time to consider a role for MRI before prostate biopsy? *Nat Rev Clin Oncol* 2009;6:197–206.
 38. Dietrich O, Heiland S, Sartor K. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magn Reson Med* 2001;45:448–453.
 39. Damon BM. Effects of image noise in muscle diffusion tensor (DT)-MRI assessed using numerical simulations. *Magn Reson Med* 2008;60:934–944.
 40. Sinha S, Sinha U. In vivo diffusion tensor imaging of the human prostate. *Magn Reson Med* 2004;52:530–537.
 41. Jambor I, Merisaari H, Taimen P, et al. Evaluation of different mathematical models for diffusion-weighted imaging of normal prostate and prostate cancer using high b-values: A repeatability study. *Magn Reson Med* 2015;73:1988–1998.
 42. Toivonen J, Merisaari H, Pesola M, et al. Mathematical models for diffusion-weighted imaging of prostate cancer using b values up to 2000 s/mm²: Correlation with Gleason score and repeatability of region of interest analysis. *Magn Reson Med* 2015;74:1116–1124.
 43. Hall MG, Bongers A, Sved P, et al. Assessment of non-Gaussian diffusion with singly and doubly stretched biexponential models of diffusion-weighted MRI (DWI) signal attenuation in prostate tissue. *NMR Biomed* 2015;28:486–495.
 44. Le JD, Tan N, Shkolyar E, et al. Multifocality and prostate cancer detection by multiparametric magnetic resonance imaging: correlation with whole-mount histopathology. *Eur Urol* 2014;67:569–576.
 45. Noguchi M, Stamey TA, McNeal JE, et al. Prognostic factors for multifocal prostate cancer in radical prostatectomy specimens: lack of significance of secondary cancers. *J Urol* 2003;170:459–463.
 46. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. London: The Cochrane Collaboration, 2011.

Godley KC, Syer T, Smith TO, Toms AT, Malcolm PN. Accuracy of high b-value diffusion-weighted MRI for prostate cancer detection: a meta-analysis. Oral presentation at Royal College of Radiologists Annual Scientific Meeting. London. September 2016

Syer TJ, Godley KC, Smith TO, Malcolm PN. Diagnostic accuracy meta-analysis of diffusion and T2 weighted imaging for prostate cancer detection: which b-value is most accurate. Oral presentation at European Congress of radiology. Vienna. March 2017.

List of formal teaching and courses attended

List of courses and conferences attended

European Congress of Radiology, Vienna, March 1-5th 2017

Royal College of Radiologists Annual Scientific Meeting, London, September 12-14th 2016

East Anglian Radiological Society, Norwich, 26th May 2016.

GCP Course. Good Clinical Practice (secondary care), Online, May 2015.

PPD sessions UEA

Improving your use of Word 2015/16

Practical Use of SPSS 2015/16

Having an Impact at a Conference 2015/16

Improving your use of Excel 2016/17

Preparing for your Viva 2016/17

Appendix 2

Radiologist reporting instructions

Instructions for Radiologists

Overview

- Four reporting sessions – cases split into batches of 20 cases.
- At least two weeks between each session
- 20 cases
- 2 reads in each session.

Case identifiers

- All patients' *Hospital number* will be **FODIP**.
- *Accession numbers* will give you access to the different cases.
- In session 1 – Accession numbers will be **1A** and **1B** etc.
- In session 2 – Accession numbers will be **1C** and **1D** etc
- **1A** and **1B** will be the same patient etc.
- **1A** and **1B** will not be the same case as **1C** and **1D**– they will be changed to a random order.
- Please report in chronological order – **1 to 20** and **21 to 40**.

Augmentation of images

Each reader can change windows centre and level to their own preference.

Template

- A reporting template will be provided for every read.
- By the end of the study each case will have 4 completed templates by each reader.

Session 1 (conventional DWI – cDWI)

- All cases will have a T1 sequence.
- The 'A' cases will contain T1 and cDWI (b100, b1000, b1500) and it's corresponding ADC map.
- The 'B' cases will contain T1, cDWI (b100, b1000, b1500) and it's corresponding ADC map, as well as a Hi-Res Axial T2 sequence
- For each case you will read **A** followed by **B** and then move onto the next case
- For example - **12A** then **12B** then **13A** then **13B**.

Session 2 (small field-of-view DWI – sDWI)

- All cases will have a T1 sequence
- The 'C' cases will contain T1 and sDWI (b100, b1000, b1500) and it's corresponding ADC map.
- The 'D' cases will contain T1, sDWI (b100, b1000, b1500) and it's corresponding ADC map, as well as a Hi-Res Axial T2 sequence
- For each case you will read **C** followed by **D** and then move onto the next case
- For example,
12C then **12D** then **13C** then **13D**.

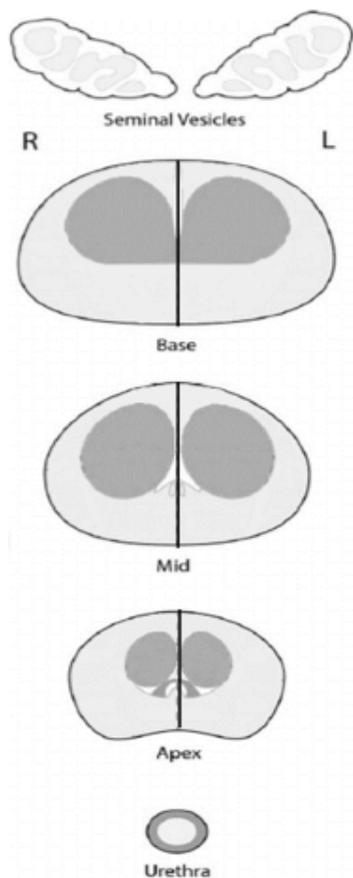
Completing the template

Patient identification

Case Number _____ **Reader** _____

Insert the Case number and reader into the specified box.

Prostate map



The prostate map is displayed here. The prostate is divided into 12 sectors.

- **Base, Mid, Apex**
- **Peripheral zone (light grey) and transitional zone (dark grey)**
- **Right and left**

Each reader decides the boundaries of the base, mid and apical gland after taking into account the size of the prostate.

The seminal vesicles are visualised also.

The outer line defining the prostate equates to the prostate capsule.

Haemorrhage, lesion location and shape, and Likert score for each sector should be documented on the map.

Prostate width – B and D cases

Maximum prostate width on T2(mm) & slice:

On T2 measure the maximum width of the prostate in mm and note the slice number.

Haemorrhage – A and C cases

Haemorrhage: Y / N

Review the T1 image and decide if haemorrhage is present and delete the below field as appropriate.

The locations of the haemorrhage should be documented on the **Prostate Map** and a '**H**' placed in the circle.

Assessment for tumour

There are two components to this

- **Lesion analysis** – all lesions suspected of being tumour should be drawn on the **Prostate Map** and characteristics drawn in the **Lesion Table**.
- **Sector analysis** – the likelihood of tumour in each sector of the prostate should be documented on the **Prostate Map**.

A Likert score will be documented for each lesion and sector.

The Likert score is as follows

- 1: Clinically significant disease is highly unlikely to be present
- 2: Clinically significant cancer is unlikely to be present
- 3: Clinically significant cancer is equivocal
- 4: Clinically significant cancer is likely to be present
- 5: Clinically significant cancer is highly likely to be present.

1. Sector analysis

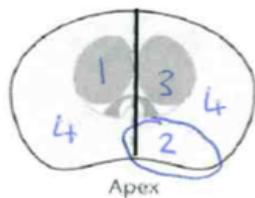
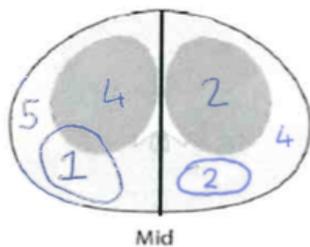
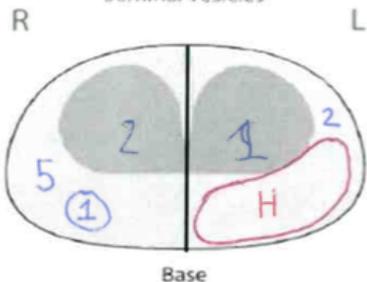
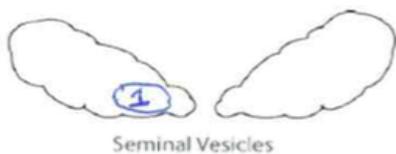
For all 12 sectors of the prostate a Likert score should be written on the **Prostate map**. This can be seen in the example below.

2. Lesion analysis

Firstly, each lesion should be drawn on the **Prostate Map** and numbered. The drawing of the tumour should as closely as possible resemble the perceived shape of the tumour. If a lesion extends from one sector to another, such as mid-gland to apex, the lesion should be drawn in each sector it resides. If the lesion extends beyond the capsule or into the seminal vesicles this should also be drawn.

In the **Lesion Table** in the corresponding Lesion number row the following should be documented.

- *Location*
 - *Base/Mid/Apex* – if in multiple sectors document this and underline the predominant place it resides.
 - *Right/Left* – if in multiple sectors document this and underline the predominant place it resides.
 - *PZ/TZ* – if in multiple sectors document this and underline the predominant place it resides.
- *Maximum axial diameter*
- *Series and slice number* – on which the maximum axial diameter was measured.
 - For cases in **A** and **C** this should be the ADC map.
 - For cases **B** and **D** this should be both the ADC map or T2 sequence.
- *Width* – perpendicular to the maximum axial diameter on the same slice.
- *Height* – The most cranial and caudal slice the tumour is visible on.
- *Likert score* – 1–5
- *ECE* – Extra-capsular extension - Yes or No – **Also document the distance of extra-capsular extension.**
- *SVI* - Seminal vesicle invasion – Yes or No.



Case Number FODIP Reader XX

Haemorrhage: Y/N

Lesion	Location			Max Axial Diameter (mm)	Series/Slice No.	Width (mm)	Height (slice)		Likert (1-5)	ECE (Y/N)	SVI (Y/N)
	B/M/A	R/L	PZ/TZ				Top	Bottom			
1	BM	R	PZ TZ	16	1100 8	8	6	9	5	N	Y
2	AM	RL	PZ	14	1100 3	6	2	4	4	Y 3mm	N
3											
4											
5											

Image Quality	1	2	3	4
Distortion (please tick)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Artefact (please tick)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Histopathology reporting instructions

FODIP

Histopathology reporting instructions

Notification of cases to histopathologists

KG will send a list of patients to RB.

Patient ID and corresponding study ID will be provided.

RB and trainee to independently review prostatectomy samples. Samples cannot be anonymised to histopathologists (ethical approval for this).

Histopathologists blinded to clinical information and MRI results.

Tumours to include

Tumours ≥ 5 mm in maximum axial diameter.

Prostate information

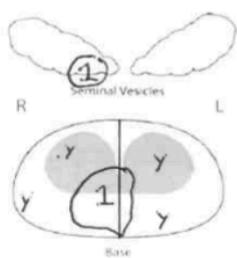
Document maximum axial diameter of prostate.

Estimated slice thickness

Lesion information

Location

- Base/middle/apex
- Right/left
- Peripheral zone/ transitional zone.
- If a lesion is in more than one sector then document all sectors involved but underline the predominant sector.
- eg



Patient Number FODIP_test1 Reader _____

Lesion	Site			Max Axial Diameter (mm)	Width (mm)	Height (mm)	ECE (Y/N)	SVI (Y/N)	Gleason
	B/M/A	R/L	PZ/TZ						
1	B	R, L	<u>PZ/TZ</u>				N	Y	
2									

Dimensions of lesion

- Maximum axial diameter (length) (mm).
- Width – perpendicular to length (mm).
- Height – (number of slices + 1) x slice thickness

Extra-capsular extension

- Y/N
- mm of extension

Seminal vesicle invasion

- Y/N

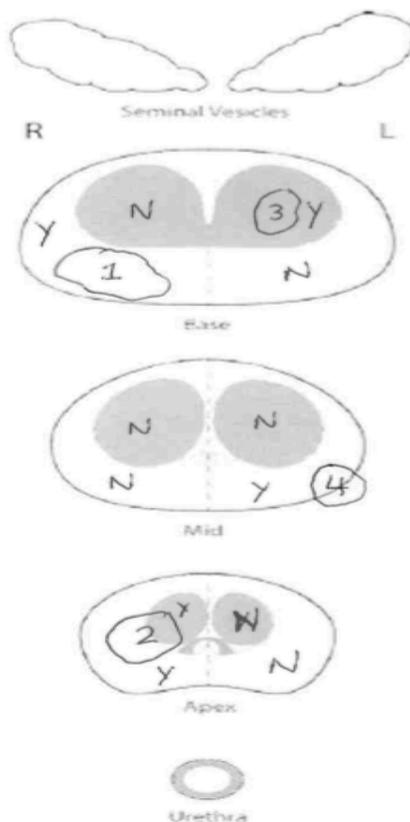
Gleason and Grade Group Score.

- Of every lesion.

Sector information

Tumour presence in each of the 12 sectors of the prostate - Y/N

eg



Following review of cases by histopathologists

- RB and trainee to provide a consensus histopathology assessment on the reporting template.
- Tumours to be outlined on each slide and a midline sagittal line drawn on the slide.
- Colour photocopies of the slides will be taken and anonymised.
- Cases reviewed with histopathologists and KG together in order to draw Regions Of Interest on the scan which as accurately as possible correspond to the exact locations of tumours.

Appendix 3

Favourable opinion of proportionate review at Research and Ethics Committee subject to conditions – 18th January 2016



18 January 2016

Dr Paul Malcolm
Consultant Radiologist
Norfolk and Norwich University Hospital
Radiology Department
Norfolk and Norwich University Hospital
Colney Lane, Norwich
NR4 7UY

Dear Dr Malcolm

Study title: Focused Diffusion Weighted Imaging in Prostate Cancer
REC reference: 16/LO/0115
IRAS project ID: 182283

The Proportionate Review Sub-committee of the London - Surrey Borders Research Ethics Committee reviewed the above application on 13 January 2016.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Mrs. Alison O'Kane, nrescommittee.london-surreyboundaries@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

The PR Sub-Committee gave a favourable opinion of the application with the following additional conditions)

1. Question A-62 in the IRAS form – please clarify the sentence 'It is not possible to blind the hisopathologists to patient name and unique hospital number.
2. The PR Sub-Committee noted a typo in A6-2 of the IRAS form, third paragraph down. The wording 'not be available' should read 'will be available' (to avoid a double negative).

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at <http://www.rdforum.nhs.uk>.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with

before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

Approved documents

The documents reviewed and approved were:

<i>Document</i>	<i>Version</i>	<i>Date</i>
IRAS Checklist XML [Checklist_21122015]		21 December 2015
Letter from funder [Pump priming grant funding]	1	30 September 2015
Letter from sponsor [FODIP NNUH sponsorship]	1	10 December 2015
Other [FODIP RCF funding]	1	30 April 2015
REC Application Form [REC_Form_21122015]		21 December 2015
Research protocol or project proposal [FODIP study protocol]	3.3	25 November 2015
Summary CV for Chief Investigator (CI) [P Malcolm CV IRAS Dec 2015]		04 December 2015
Summary CV for student [K Godley IRAS CV]		
Summary CV for supervisor (student research) [A Toms IRAS CV]		02 December 2015

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:

<http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/>

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at <http://www.hra.nhs.uk/hra-training/>

With the Committee's best wishes for the success of this project.

16/LO/0115

Please quote this number on all correspondence

Yours sincerely



pp
Sir Adrian Baillie
Chair

Email: nrescommittee.london-surreybounders@nhs.net

Enclosures: List of names and professions of members who took part in the review

"After ethical review – guidance for researchers"

*Copy to: Dr Keith Godley
Mrs Karen Baucutt, Norfolk and Norwich University Hospital*

Favourable opinion of proportionate review at Research and Ethics Committee

– 22nd January 2016



22 January 2016

Dr Paul Malcolm
 Consultant Radiologist
 Norfolk and Norwich University Hospital
 Radiology Department
 Norfolk and Norwich University Hospital
 Colney Lane, Norwich
 NR4 7UY

Dear Dr Malcolm

Study title: Focused Diffusion Weighted Imaging in Prostate Cancer
REC reference: 16/LO/0115
IRAS project ID: 182283

Thank you for your email of 21st January 2015. I can confirm the REC has received the documents listed below and that these comply with the approval conditions detailed in our letter dated 18 January 2016

Documents received

The documents received were as follows:

Approved documents

The final list of approved documentation for the study is therefore as follows:

<i>Document</i>	<i>Version</i>	<i>Date</i>
IRAS Checklist XML [Checklist_21122015]		21 December 2015
Letter from funder [Pump priming grant funding]	1	30 September 2015
Letter from sponsor [FODIP NNUH sponsorship]	1	10 December 2015
Other [FODIP RCF funding]	1	30 April 2015

Other [email from researcher following favourable with conditions decision]		21 January 2016
REC Application Form [REC_Form_21122015]		21 December 2015
Research protocol or project proposal [FODIP study protocol]	3.3	25 November 2015
Summary CV for Chief Investigator (CI) [P Malcolm CV IRAS Dec 2015]		04 December 2015
Summary CV for student [K Godley IRAS CV]		
Summary CV for supervisor (student research) [A Toms IRAS CV]		02 December 2015

You should ensure that the sponsor has a copy of the final documentation for the study. It is the sponsor's responsibility to ensure that the documentation is made available to R&D offices at all participating sites.

16/LO/0115	Please quote this number on all correspondence
-------------------	---

Yours sincerely

Alison O'Kane

Alison O'Kane
Acting REC Manager

E-mail: nrescommittee.london-surreyborders@nhs.net

Copy to: *Dr Paul Malcolm, Norfolk and Norwich University Hospital*
Dr Keith Godley
Mrs Karen Baucutt, Norfolk and Norwich University Hospital

Non-substantial amendment approval – 15th August 2016

RE: (IRAS 182283. Confirmation of Amendment Categorisation as Category C)

amendments hra (HEALTH RESEARCH AUTHORITY) [hra.amendmen...

To: [GODLEY KEITH \(RM1\) Norfolk and Norwich University Hospital](#)

Cc: [MALCOLM PAUL \(RM1\) Norfolk and Norwich University Hospital](#)

28 July 2016 10:29

- You forwarded this message on 15/08/2016 14:43.

Dear Keith Godley,

Further to the below, I am pleased to confirm that HRA Approval has been issued for the referenced amendment, following assessment against the HRA criteria and standards.

The sponsor should now work collaboratively with participating NHS organisations in England to implement the amendment as per the below categorisation information. This email may be provided by the sponsor to participating organisations in England to evidence that the amendment has HRA Approval.

Please contact hra.amendments@nhs.net for any queries relating to the assessment of this amendment.

Yours sincerely,

Dr Claire Cole
Senior Assessor
