Evaluation of an MRI-based screening pathway for prostate cancer

A thesis presented for the degree of

Doctor of Philosophy (PhD)

David Eldred-Evans MA MBBS

Imperial Prostate

Department of Surgery and Cancer

Imperial College London

Copyright Declaration

Abstract

In recent years there has been a wealth of debate regarding prostate cancer screening, with a concurrent increase in new imaging techniques for prostate cancer diagnosis. Imaging has been the technique of choice in lung and breast cancer screening programmes but has not been explored for prostate cancer screening. Herein, this thesis explores the role of magnetic resonance imaging (MRI) as a new approach to screen for prostate cancer.

Following an introduction to the current screening landscape, my thesis focuses on the development and validation of a fast MRI, known as a prostagram, that could serve as a viable image-based screening test. Evaluation of this new technique is performed within a prospective, population-based, blinded, cohort study which was conducted at seven primary care practices and two imaging centres. A diverse array of performance characteristics of fast MRI are compared to PSA. These encompass biopsy rates, cancer detection rates, diagnostic accuracy and patient reported experience measures.

The second half of this thesis focuses on further optimising the fast MRI protocol for screening and exploring methods of integrating it into an alternative screening pathway. The outcomes point towards a pathway which combines a low threshold PSA and a fast MRI as yielding a more acceptable balance between benefits and harms. This is followed by the development of a risk tool to address the challenges of equivocal MRI lesions.

Overall my thesis provides a balanced evaluation of fast MRI as a new screening test and the final chapter highlights outstanding challenges that must be addressed for fast MRI to progress as a legitimate screening modality. There is a requirement for all new screening tests to be evaluated in robust randomised controlled trials and the thesis concludes by setting out a phased research framework for fast MRI to enable a full evaluation over the next decade.

Declaration of Originality

I hereby declare that the work presented in this thesis was undertaken by me. This thesis, and the work described therein was planned and carried out over a three year period as a Clinical Research Fellow at Imperial Prostate, Imperial College London. Collaboration and important input from other individuals is explicitly referenced in the respective chapters.

Acknowledgements

This thesis would not have been possible without the guidance and support of several individuals who I have been privileged to work with. I am, and always will be, deeply indebted to my primary supervisor, **Professor Hashim U. Ahmed**, who guided and advised me throughout my research. Hash has been far more than an academic supervisor – he has been an inspirational mentor, a trusted surgical tutor and a valued friend. During the course of the years, Hash has given me unprecedented freedom to explore topics of my interest and empowered me to take a research idea from funding applications through to implementation in a clinical trial.

I owe an enormous debt of gratitude to **Mr Mathias Winkler** who took a chance initially appointing me as a research fellow at Imperial. He has been a constant source of encouragement and his knowledge, understanding and experience of prostate cancer is unsurpassed. He leads by example and I will always be thankful to him for instilling me with the confidence that is needed to develop and then deliver large and multifaceted projects such as the IP1-PROSTAGRAM trial and RAPID Online.

Dr Henry Tam was my starting point for designing the fast MRI or 'prostagram' and has been attached to these projects since the beginning. Thank you for allowing me into the world of radiology and for the hundreds of hours you spent reporting the prostagrams. Supervising an imaging-driven thesis in prostate cancer did not come without a certain added risk to all my supervisors. I am grateful to Hash, Matt and Henry for going beyond the call of duty and volunteering to be scanned as subjects in 2017 when I was developing the prostagram MRI protocol.

This thesis has culminated in many enriching collaborations and there are many more people who have helped in different aspects than I can individually thank here. From Paul Strickland Scanning Centre, I would like to offer my heartfelt thanks and appreciation to **Dr Hemmy Sokhi** who helped drive through the execution of the trial and invested a considerable amount of time reporting many hundreds of scans. Particular thanks are due to **Professor Anwar Padhani** for sharing his immense depth of experience in prostate MRI and undertaking to report as a second radiologist.

From University Medical Center Utrecht, I would like to thank **Dr Max Peters** from whom I have learnt so much about statistics, modelling and data science. Before starting this PhD I

had not written a single line of code or indeed heard of R Statistic Software. His encouragement and experience spurred me on to learn the language and move towards becoming a proficient R programmer.

At Imperial, this work would not have been completed if it was not for the help and assistance received from Feargus Hosking-Jervis, Paula Burak, Martin Evans and Emma Cullen. **Feargus** has been a valued source of friendship, support and good humour since our first day at Imperial Prostate. **Paula's** immense organizational abilities helped me navigate the complex task of regulatory approvals and her ongoing efforts kept the study progressing on time and on budget. **Martin** proved to be the linchpin of navigating men through the trial and was willing to turn his hand to almost any task or challenge we faced while always being upbeat. **Emma** was a source of support in all senses of the word and had a gift for always providing practical, positive assistance whenever needed.

Any successes of this thesis are fundamentally dependent on the willingness of men to volunteer for research and I am immensely grateful to over a thousand men who participated across the studies used in this thesis. Without their willingness to contribute to medical research none of my projects could have proceeded.

With regard to the IP1-PROSTAGRAM study I would like to acknowledge the help of many team members. I could not have completed the study without the input of Martin Connor, Lorraine O'Connell, Kerry Richards, Ivana Russo, Albert Busza, Emily Day, Francesca Fiorentino, Natalia Klimowska-Nassar, Sandra O'Sullivan, Marta Zerominska, Patrick Daly, Janki Bhayani, Fatima Akbar, Ross Dalton-Short, Mariana Bertoncelli, Mohamed Noureldin, Christopher Khoo, Taimur Shah, Saiful Miah, Deepika Reddy, Nishant Bedi, Derek Lomas, Arnas Rakauskas and Heather Bhola-Stewart.

Special thanks to my family; my father-in-law, **Malcolm**, who was willing to spend many days travelling to different sites for repeated prostate MRIs to optimise the scanners. My mother-in-law, **Caroline**, for giving me an alternative place to write-up as everywhere locked-down. My father, **Charles**, for proof reading the many pages of this thesis with such attention to detail. My mother, **Susan**, for supporting me throughout my life. My son, **Joseph**, who was born during this PhD and has been a fantastic source of pure fun.

And finally, to my wife **Ami** who has made many sacrifices to help me through this process, indeed anything I have achieved was and is only possible because of her.

Funding

The author is grateful for the following sources of funding towards this PhD:

- A BMA Foundation for Medical Research grant: £100,000 (Eldred-Evans)
- A Urology Foundation Research Scholarship: £59,192 (Eldred-Evans)
- Imperial Health Charity Research Fellowship: £50,000 (Eldred-Evans)
- NIHR Imperial Biomedical Research Centre Grant: £15,000 (Ahmed & Eldred-Evans)
- A Royal College of Surgeons of England Honorary Research Fellowship: £3,000 (Eldred-Evans)

The majority of the work related to the IP1-PROSTAGRAM trial was funded by a Wellcome Trust Programme Senior Clinical Fellowship Grant awarded to Professor Ahmed (Grant number: 204998/Z/16/Z)

The BMA Foundation for Medical Research grant supported the work in Chapters 5-8. The Urology Foundation Research Scholarship supported the work in Chapter 9. The Imperial Health Charity Research Fellowship supported the work in Chapters 11-12. The NIHR Imperial Biomedical Research Centre grant supported the work in Chapter 6. The Royal College of Surgeons of England Honorary Research Fellowship supported the work in Chapters 2-3.

The views and opinions expressed in this thesis are those of the candidate and do not necessarily reflect those of the NHS, the NIHR, the BMA Foundation, Imperial Health Charity, the Royal College of Surgeons or the Wellcome Trust.

List of publications and presentations arising from this thesis

Publications

Original Research

Eldred-Evans D, et al Population-based prostate cancer screening with Magnetic Resonance or Ultrasound Imaging: The IP1-PROSTAGRAM study. 2021. *JAMA Onc (In Press)*.

Eldred-Evans D, et al. Added value of diffusion-weighted images and dynamic contrast enhancement in multiparametric magnetic resonance imaging for the detection of clinically significant prostate cancer in the PICTURE trial. *BJU Int*. 2020 Mar;125(3):391-398. doi: 10.1111/bju.14953. Epub 2019 Dec 11. PMID: 31733173.

Khoo, C.C., **Eldred-Evans, D.**, Peters, M., Bertoncelli Tanaka, M., ... & Ahmed, H.U i,. Likert vs PI-RADS v2: a comparison of two radiological scoring systems for detection of clinically significant prostate cancer. *BJU International*, 2019. 125(1), pp.49-55. PMID: 31599113

Khoo, C.C., **Eldred-Evans, D.,** Peters, M., van Son, M., ... & Ahmed, H.U "A Comparison of Prostate Cancer Detection Between Visual-Estimation (Cognitive Registration) and Image-Fusion (Software Registration) Targeted Transperineal Prostate Biopsy." The Journal of Urology (2020): 10-1097. PMID: 33207137

Miah S, **Eldred-Evans D**, Simmons LA, Shah TT, Kanthabalan A, Arya M, ... & Ahmed, H.U. Patient reported outcome measures for transperineal template prostate mapping biopsies in the PICTURE study. *The Journal of urology*. 2018 Dec 1;200(6):1235-40. PMID: 29940251

Miah, S., Hosking-Jervis, F., Connor, M. J., **Eldred-Evans, D**., ... & Ahmed, H.U. A multicentre analysis of the detection of clinically significant prostate cancer following transperineal image-fusion targeted and nontargeted systematic prostate biopsy in men at risk." European urology oncology 3.3 (2020): 262-269. PMID: 31411968

Reviews

Eldred-Evans D, et al. Rethinking prostate cancer screening: could MRI be an alternative screening test?." *Nature Reviews Urology* 17.9 (2020): 526-539. PMID: 32694594

Eldred-Evans, D, et al. Use of imaging to optimise prostate cancer tumour volume assessment for focal therapy planning, *Current Urology Reports* 21.10 (2020): 1-10. PMID: 30821096.

Connor, M.J., Miah, S., Jayadevan, R., Khoo, C.C., **Eldred-Evans, D**., Shah, T., Ahmed, H.U. and Marks, L., Value of systematic sampling in an mp-MRI targeted prostate biopsy strategy. *Transl Androl Urol*. 2020. PMID: 32676437

Lovegrove, C.E., Matanhelia, M., Randeva, J., **Eldred-Evans, D.**, Tam, H., Miah, S., Winkler, M., Ahmed, H.U. and Shah, T.T., 2018. Prostate imaging features that indicate benign or malignant pathology on biopsy. *Transl Androl Urology*, 7(Suppl 4), p.S420. PMID: 30363462

Noureldin M, **Eldred-Evans D**, Khoo CC, Winkler M, Sokhi H, Tam H, Ahmed HU. MRI-targeted biopsies for prostate cancer diagnosis and management. *World J Urol*. 2020 Apr 6. doi: 10.1007/s00345-020-03182-3. Epub ahead of print. PMID: 32253585.

Commentary and Letters

Eldred-Evans D, Ahmed HU. Re-thinking active surveillance for the multiparametric magnetic resonance imaging era. *BJU Int*. 2019 Mar;123(3):376-377. doi: 10.1111/bju.14699. Erratum in: BJU Int. 2019 May;123(5):909. PMID: 30821096.

Eldred-Evans D. Winkler M., Ahmed HU.. MRI screening for prostate cancer: a step towards a 'prostagram'. *Urology News* 2020 Jan, Vol 24 No. 2

Jambor, I., Falagario, U, Martini, A., **Eldred-Evans, D**. et al. Re: Variability of the Positive Predictive Value of PI-RADS for Prostate MRI across 26 Centers: Experience of the Society of Abdominal Radiology Prostate Cancer Disease-focused Panel. European urology OnCology, 78(4), pp.633-636.

Connor, M.J, Gorin M, **Eldred-Evans, D**. Desai A, … & Ahmed, H.U. Landmark Developments in the Evolution of the Modern Prostate Biopsy Procedure: A Look Back and Forwards

Oral Presentations with published Abstracts

Eldred-Evans, D., Burak, P., Connor, M.J. et al. Population-based prostate cancer screening using a prospective, blinded, paired screen-positive comparison of PSA and fast MRI: The IP1-PROSTAGRAM study. *Journal of Clinical Oncology 38*,
DOI: 10.1200/JCO.2020.38.15_suppl.5513 (Virtual meeting)

Eldred-Evans, D, Joshi S., Pegers E., Ahmed, HU,. Rapid Access to Prostate Imaging and Diagnosis (RAPID) *Cancer Research UK Fifth Biennial Early Diagnosis Research Conference 2019*

Poster Presentations with published abstracts

Eldred-Evans, D, Bertoncelli Tanaka, M et al. Rapid Access Prostate Imaging and Diagnosis (RAPID) pathway – an innovative approach for prostate cancer diagnosis - *Journal of Clinical Urology* 2019, Vol. 12(1S) 9–89

Khoo, C. **Eldred-Evans, D**, et al. Man vs. Machine: Comparing Cognitive and Software-Assisted mpMRI-Ultrasound Fusion Targeted Biopsy- *Journal of Clinical Urology* 2019, Vol. 12(1S) 9–89

Connor C., Miah, S. **Eldred-Evans D** et al. Crossing the perineum: history of transperineal prostate biopsy, *Journal of Clinical Urology* 2019, Vol. 12(1S) 9–89

Noureldin, M, **Eldred-Evans D**, et al. Implementation of MDT evaluating multi-parametric prostate MRI: Is double reporting needed? Mr Mohamed Noureldin. *Journal of Clinical Urology* 2019, Vol. 12(1S) 9–89

Bertoncelli Tanaka, M, **Eldred-Evans D**, Re-thinking the prostate cancer diagnostic pathway: the acceptability of the Rapid Access Prostate Imaging and Diagnosis (RAPID) pathway Mr Mohamed Noureldin. *Journal of Clinical Urology* 2019, Vol. 12(1S) 9–89

Eldred-Evans, D, Peters M., Connor M., et al The Rapid Risk Score: Development of a Novel Risk Score to Predict Significant Prostate Cancer in men with an mpMRI lesion" *The Journal of Urology* 203.Supplement 4 (2020): e851-e852.

Eldred-Evans, D., Burak, P., Connor, M.J. et al. Population-based prostate cancer screening using a prospective, blinded, paired screen-positive comparison of PSA and fast MRI, *The Journal of Urology*, 203(Supplement 4), e1115-e1115.

Bass, E., **Eldred-Evans, D**., Connor, M. et al, The RAPID Access Prostate Imaging and Diagnosis (RAPID) Pathway – A multicentre update of 1719 patients undergoing multi-parametric MRI as a triage test, *The Journal of Urology*, 203(Supplement 4), pp.e186-e186.

Khoo, C., **Eldred-Evans D**., Peters M., et al, Man vs. Machine: Comparative Effectiveness of Cognitive Targeted and Image-Fusion Targeted Transperineal Prostate biopsy, *The Journal of Urology*, 203(Supplement 4), pp.e847-e848.

Connor, M.J., **Eldred-Evans, D**., Feargus, H.J., et al Direct and Marginal Cost Analysis of not aiming for the Target in a MRI-Targeted Prostate biopsy pathway. *The Journal of Urology*, 203(Supplement 4), pp.e413-e414.

Khoo, C.C., **Eldred-Evans, D**., Jaenicke, J. et al Likert vs. PI-RADS v2: Comparison of two radiological scoring systems for detection of clinically significant prostate cancer. *The Journal of Urology*, 201(Supplement 4), pp.e1096-e1096.

Connor, M.J., **Eldred-Evans, D**., van Son, M., et al . A Multicentre Study of the Clinical Utility of Non-targeted Systematic Transperineal Prostate Biopsies in Patients Undergoing Pre-biopsy mpMRI. *The Journal of Urology*, pp.10-1097.

Khoo C.C. **Eldred-Evans D**. Jaenicke J.et al. Likert vs. PI-RADS v2: a comparison of two radiological scoring systems for detection of clinically significant prostate cancer. *Eur Urol Suppl*. March 2019; 18: e1865-e1866

Eldred-Evans D., Brittain J., Servian P., et al. The role of multi-parametric MRI as a triage test: A propensity-matched comparison of a MRI-triage and a TRUS-biopsy pathway. *European Urology Supplements*, Volume 17, Issue 2, March 2018; Pages e695-e696

Maynard W. **Eldred-Evans D.,** Connor M. et al Local anaesthetic transperineal prostate biopsy: Optimising patient selection. *European Urology Open Science* 19 (2020): e2319.

Khoo C.C, **Eldred-Evans D.,** Peters M. et al, Man vs machine: Comparative effectiveness of cognitive targeted and image-fusion targeted transperineal prostate biopsy. *European Urology Open Science* 19 (2020): e1413

Reddy D.T. **Eldred-Evans D.,** Connor M. J. et al, Indeterminate mpMRI lesions: Evaluating the optimal PSA density threshold for prostate biopsy. *European Urology Open Science 19 (2020): e1702-e1703*

Connor, M.J., **Eldred-Evans, D**., Hosking-Jervis. F.., et al, Direct and Marginal Cost Analysis of not aiming for the Target in a MRI-Targeted Prostate biopsy pathway. *European Urology Open Science 19 (2020): e1073-e1074*

Connor, M.J., **Eldred-Evans, D**., Hosking-Jervis. F..,., et al Which men should undergo non-targeted systematic sampling in an mpMRI-targeted pathway – an analysis from 1,719 pre-biopsy mpMRI cases?. *European Urology Open Science 19 (2020): e191.*

Eldred-Evans D., Peters M, Bertoncelli Tanaka M et a. The RAPID risk model: A novel risk score to predict significant prostate cancer in men with an mpMRI lesion. *European Urology Open Science* 19 (2020): e1741-e1742.

Rakauskas A. *, Shah T., Hosking-Jervis F, **Eldred-Evans D** et al. Transperineal biopsy of the prostate with image fusion: A suggested technique and cancer detection rates: *European Urology Open Science 19 (2020): e2341.*

Invited presentations

1. Getting it right: Indications for modern urological imaging, Can Transperineal Biopsies Solve the problem. *ESUI18: 7th Meeting of the EAU Section of Urological Imaging*, Amsterdam (2018)

2. Diagnosis in Prostate Cancer Debate: The Case for MRI. *The North of England Urological Society*, Newcastle (2019)

Prizes and Awards

1. TUF Medal for best research application (2018)

2. TP Gunton Award, BMA Foundation (2018)

3. Helen H Lawson Award, BMA Foundation (2018)

4. Chair Award's (Finalist), Imperial College NHS Trust (2020)

Media Coverage and Public Engagement

Considerable media coverage was received in both national and international press related to research undertaken as part of this PhD. Selected examples are:

Radio

- BBC Health Check, Dr Anne Robinson. Segment on "A new way of screening for prostate cancer". 20th May 2020 Available at www.bbc.co.uk/programmes/w3cszcbt

- LBC, Nick Ferrari Breakfast Show. "Reasons to be cheerful". 22nd May 2020 Available from https://www.globalplayer.com/catchup/lbc/uk/46vyD7z/

- UCB, Interview of D Eldred-Evans by Paul Hammond "Talking Point". 14th May 2020

National Newspapers

- The Times, "Breakthrough in prostate cancer screening". 14th May 2020.

- The Telegraph, 14th May 2020. Available from https://www.telegraph.co.uk/news/2020/05/14/game-changing-prostogram-takes-fear-ofprostatecancer-testing/

- BBC Science Focus: 14th May 2020. Available from https://www.sciencefocus.com/news/game-changing-prostate-cancer-test-works-in-15-minutes/

- The Daily Mail, 14th May 2020. Available from www.dailymail.co.uk/news/article-8317461/Screening-middle-aged-men-prostate-cancer-pick-8-000-extra-cases.html

- Daily Express, 14th May 2020. Available from https://www.express.co.uk/life-style/health/1282249/Superscan-screening-prostate-cancer-latest-health-news

- The Mirror, 14th May 2020. Available from https://www.mirror.co.uk/news/uk-news/game-changing-cancer-scan-created-22024465

- The Sun, 14th May 2020. Available from https://www.thesun.co.uk/news/health-news/11615602/quick-15-minute-scan-could-save-thousands-of-lives-detecting-prostate-cancer-early/

International Newspapers

- Spain Today, 14th May 2020. Available from https://spaintoday.es/uncategorized/british-scientists-announce-breakthrough-prostate-cancer-scan-that-gives-results-in-just-15-minutes/

- Euro Weekly News, 14th May 2020. Available from https://www.euroweeklynews.com/2020/05/14/british-scientists-announce-breakthrough-prostate-cancer-scan-that-gives-results-in-just-15-minutes/

- Gilbraltar Chronicle, 14th May 2020. Available from https://www.chronicle.gi/scientists-develop-game-changing-15-minute-prostate-cancer-test/

- FR24 News, 14th May 2020. Available from https://www.fr24news.com/fr/a/2020/05/une-analyse-rapide-de-15-minutes-pourrait-sauver-des-milliers-de-vies-detecter-tot-le-cancer-de-la-prostate-the-sun.html

- Irish Examiner, 14th May 2020. Available from https://www.irishexaminer.com/breakingnews/world/scientists-develop-game-changing-15-minute-prostate-cancer-test-999451.html

- Thakon News, 14th May 2020. Available from https://thakoni.com/scientists-develop-game-changing-test-for-prostate-cancer/

- News Colony, 14th May 2020. Available from https://newscolony.com/superscan-could-be-the-key-to-screening-for-prostate-cancer/

- Lifestyle to the minute, 14th May 2020. Available from https://www.noticiasaominuto.com/lifestyle/1478693/muda-as-regras-do-jogo-teste-de-15-min-pode-salvar-milhares-de-vidas

Other Media

- Prostate Cancer UK. "Behind the headlines: 'game-changing' new scan could pave the way to screening". Available from https://prostatecanceruk.org/about-us/news-and-views/2020/5/game-changing-new-scan-could-pave-the-way-to-screening

- Wide Coverage in UK Regional Newspapers: Evening Standard, Express & Star, Wales Online, Evening Express, Herefordshire Mercury, The Reading Evening Post, Berkshire Live, Grimsby Telegraph, Kentlive, Bournemouth Echo, Essex Live, Aberdeen Evening Express, Wiltshire Times, Tottenham Independent, Glasgow Times, Basingstoke Gazette, Bury Times, Swindon Advertiser, Internewscast, Jersey Evening Post, Derby Evening Telegraph, Guernsey Press, Ayr Advertiser, Salford City News, Cambridge Evening News, Belfast Telegraph, Western Telegraph, Malvern Gazette, The National Scot, Carrick Herald, Campaign, Ashbourne News Telegraph.

Table of Contents

List of Figures

List of Tables

Chapter 1- Introduction

1.1 Rationale of thesis

"PSA is still a poor test for prostate cancer and a more specific and sensitive test is needed" (UK National Committee Screening)[1]

The UK National Screening Committee has identified a need for new screening tests for prostate cancer[1]. There is currently no national screening programme for prostate cancer and the lack of a reliable test, or combination of tests, has been recognized as a key barrier to progress towards an acceptable screening programme. This thesis aims to advance our understanding into the potential role of prostate MRI as a screening test, either as a stand-alone test or incorporating it with other tests within a new screening pathway.

Prostate cancer is a major public health burden, with 1.27 million new cases and 360,000 deaths worldwide[2]. Though the mortality rate is falling slowly due to early detection and improved treatments, it remains a leading cause of death in most developed countries[3]. The lifetime risk of dying from prostate cancer among men is 4.3%[4] and the number of deaths have recently overtaken those from breast cancer in the UK.

The aim of a screening programme would be to detect life-threatening prostate cancer at a curable stage and thereby reduce cancer-specific mortality. Supporters of screening highlight that prostate cancer is an anomaly amongst common cancers as unlike breast and colorectal cancer, it lacks an established screening programme.

Meanwhile, the shift in age composition of the population will make screening and early detection of prostate cancer a central issue over the coming decades. The number of men aged ≥50 years old globally is estimated to rise to one billion by 2030, compared to 300 million in 2000[5]. This will place significant upward pressures on prostate cancer incidence and mortality with the number of deaths from prostate cancer projected to grow from 360,000 to 740,000 by 2040[6].

On the other hand, prostate cancer poses a unique challenge due to the large reservoir of low risk or 'clinically insignificant' prostate cancer, estimated as occurring in approximately 1 in 3 men aged ≥50 years old [7]. Previous attempts to introduce prostate cancer screening using prostate specific antigen (PSA) have proven the widespread harms generated by overdiagnosis[8]. National screening committees have been consistent in reaching the

conclusion that the harms of PSA screening outweigh the benefits on a population level. Thus there is a need to re-evaluate the PSA-based screening model and consider alternative screening tests for prostate cancer.

1.2 Aims of thesis

At the start of planning the work described in this thesis, MRI was a technique which had been barely considered as a screening test. Many prior studies had evaluated multi-parametric MRI as a test in secondary care for men with a suspicion of prostate cancer where it had high diagnostic accuracy. However, these studies had not considered MRI as a screening test and it was unclear whether MRI could have the correct performance characteristics and attributes for screening.

The aim of this thesis was to address a gap in the evidence and provide new information on whether MRI has a role as a screening test for prostate cancer. The work draws on the experience of other common cancers where image-based screening tests have been successfully adopted. For instance, mammography has been an established screening programme for breast cancer and there is mounting evidence for the efficacy of low-dose CT for lung cancer[9].

The first objective of this thesis was to validate an alternative MRI technique, known as a prostagram, that could serve as a viable image-based screening test. As with other screening tests, an effective screening technique for prostate cancer should, at a minimum, be feasible, safe, accurate and acceptable. The subsequent thesis objectives were to compare fast MRI to PSA as screening tests with respect to these attributes. This involved designing a prospective, population-based, blinded cohort clinical trial, the Imperial Prostate 1 PROSTAGRAM (IP1-PROSTAGRAM) study, in which men were screened for prostate cancer using both PSA and a fast MRI.

The ensuing objectives were to understand the acceptability of fast MRI as a screening test and whether any adjunct tests could be combined with fast MRI to improve the efficiency of the pathway. The final objective was to explore practical methods of incorporating fast MRI into the whole screening pathway, incorporating the diagnostic pathway in secondary care, in combination with other clinical parameters including PSA.

1.3 Thesis outline

My thesis is comprised of 12 chapters documenting the work conducted for this research degree. Each chapter will consist of a short overview of the chapter followed by a more detailed introduction, methods, results and conclusions. The chapters are grouped into four parts as follows:

Part 1 comprises three introductory chapters which provide the context for the thesis. **Chapter 2** begins with a primer in some core concepts for screening and explores the specific challenges surrounding prostate cancer screening. In this chapter, I explore the deficiencies of the PSA-based approach to screening and provide the rationale for this new approach. The introductory section continues in **Chapter 3** with a detailed narrative review of prior work on MRI and the potential challenges to implementing MRI screening.

Part 2 explains the process of developing and evaluating a fast MRI protocol for prostate cancer screening. **Chapter 4** presents an analysis of the PICTURE trial which was a paired-cohort validating confirmatory study allowing a comparison of the diagnostic value of additional MRI sequences. Using these results a bi-parametric MRI was selected as the most appropriate screening test to be evaluated in my doctoral programme of work. **Chapter 5** describes the design and recruitment to the IP1-PROSTAGRAM Trial which was a prospective, population-based, blinded cohort study. **Chapter 6** I presents the primary outcomes of IP1-PROSTAGRAM which compares the diagnostic performance of PSA and MRI as a screening and then, in **Chapter 7**, show how the diagnostic accuracy of MRI and PSA in IP1-PROSTAGRAM can be adjusted to correct for verification bias. **Chapter 8** focuses on patient experience and compares the acceptability of fast MRI and PSA screening using patient-reported experience measures (PREMs). **Part 2** concludes that fast MRI has characteristics which make it attractive as a screening test although it may need to be combined with other tests in a screening pathway.

Part 3 includes three chapters which evaluate combining PSA and fast MRI within a new multimodal screening pathway. **Chapter 9,** compares a range of pathways and thresholds to identify a pathway which maintains diagnostic accuracy without excessive false positives or overdiagnosis. In **Chapter 10,** I evaluate the current use of MRI in a rapid access diagnostic pathway as a model which could be used to deliver this combined pathway in a one-stop setting and further, in **Chapter 11**, I build on the outcomes of this pathway to develop and validate an integrated risk prediction model to further reduce biopsy rates. These latter two

chapters evaluate strategies that could be employed across a healthcare sector to further reduce the potential harms and burdens that come from diagnosing prostate cancer.

The final section of my thesis summarises the discoveries of the thesis and brings together the broad conclusions from each chapter. In **Chapter 12** the strength and limitations of my work are discussed and the findings are placed in the wider context to provide directions for future research.

Chapter 2 – Background

2.1 Overview

This chapter sets the context for the thesis by outlining fundamental principles of screening and specific challenges of prostate cancer screening. It then outlines the current status of prostate specific antigen (PSA) as a screening test and summarises the evidence from randomised controlled trials. A portion of this chapter includes work published in Current Urology Reports[i].

2.2 Introduction

Although screening tests have a similar objective to standard diagnostic tests, aiming to detect the presence or absence of disease, there is a fundamental difference in the target population. Screening tests are performed across a large population of asymptomatic individuals, the majority of whom are healthy and do not have the target disease.

This leads to several confounding factors inherent to screening which make evaluating a new screening test prone to bias. These include,

- Length-time bias: The propensity for screening to detect slower-growing cancers which have a longer natural history and are likely to have a better outcome.

- Overdiagnosis bias: An extreme form of length-time bias where the disease detected would never have caused morbidity or mortality. Without screening these cancers would not have been detected but due to screening there is a risk of overtreatment.

- Lead-time bias: Screening can identify disease earlier in its natural history so survival time can appear longer even if the outcome remains unchanged.

[i] Eldred-Evans, David, et al. Current Urology Reports 21.10 (2020): 1-10.

Recognising these problems, Wilson and Jungner wrote the *Principles and Practices of Screening for Disease* for the World Health Organisation in which ten fundamental principles for screening were set out[10]:

- The condition sought should be an important health problem.
- The natural history of the condition, including development from latent to declared disease, should be adequately understood.
- There should be a recognizable latent or early symptomatic stage.
- There should be a suitable test or examination.
- The test should be acceptable to the population.
- There should be an agreed policy on whom to treat as patients.
- There should be an accepted treatment for patients with recognized disease.
- Facilities for diagnosis and treatment should be available.
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
- Case-finding should be a continuing process.

The relevant principles are discussed through this chapter in the context of prostate cancer screening and PSA as a screening test.

2.3 Prostate Cancer Mortality

"The condition sought should be an important health problem"
World Health Organisation [10]

Prostate cancer is a leading cause of cancer death in men. This can sometimes be underemphasised as the debate on prostate cancer screening often revolves around the challenges of overdiagnosis and overtreatment related to screening. Although important, it does not negate that prostate cancer remains the second most common cause of male cancer deaths and it accounts for 14% of all cancer deaths in males in the UK. The number of deaths has been persistently increasing each year reaching 12,032 in 2017[11].

Mortality from prostate cancer is strongly determined by the age distribution of the population. Age-standardised mortality rates are a more refined method of evaluating the burden of disease than crude total death rates. Age-standardised rates can be used to compare mortality rates over time without being affected by any changes in the age-distribution of the population. A comparison of age standardised mortality rates and the crude total number of deaths from prostate cancer are given in Figure 1.

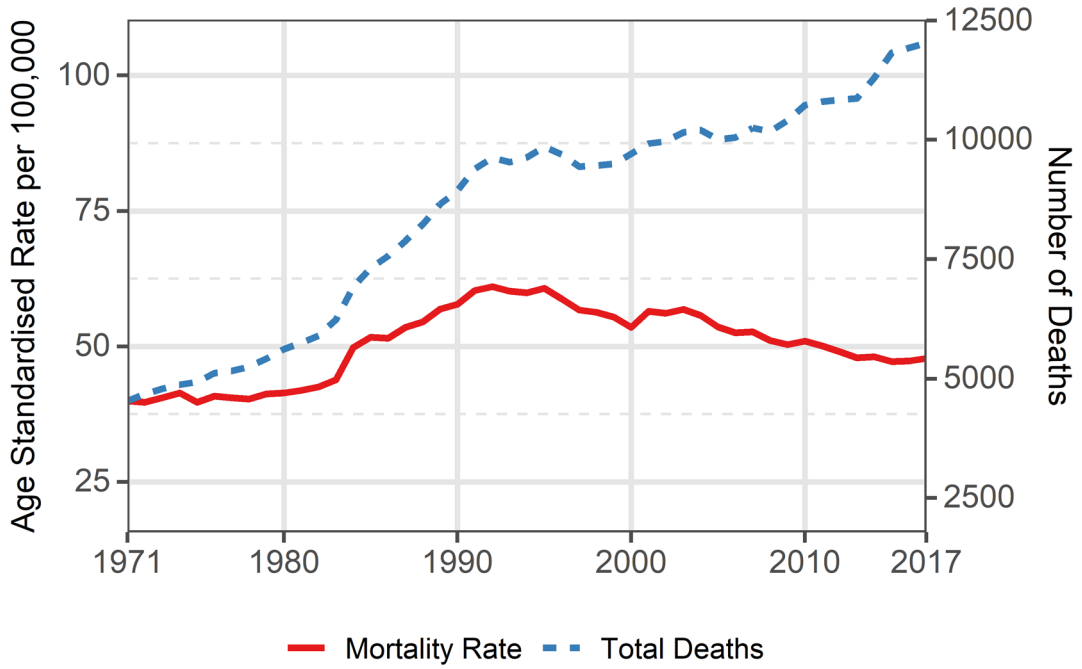


Figure 1: Trends in Age-Standardised Mortality Rates and Total Number of Deaths for prostate cancer in the UK. The age-standardised weight is a weighted average of the age-specific mortality rates per 100,000 persons. standardised using the 2013 European Standard Population. Source: CRUK[11]. The absolute number of deaths due to malignant neoplasm of the prostate (C61) includes all deaths reported by the UK to the WHO Regional Office for Europe. Source: WHO[12].

The trend was initially similar with a rapid increase in both mortality rate and total number of deaths during the 1990s. The age-standardised mortality rate has stabilised and gradually decreased in recent years while the total number of deaths has continued to increase. This trend has been similar in most developed countries across the world[13].

Only limited modifiable risk factors causing prostate cancer have been recognised[14]. This has limited the scope of primary prevention as a public health measure to reduce mortality rates and increased the need for a screening programme. Instead the difference between age-standardised and total numbers of deaths reflects an ageing population and in particular an increase in number of men aged 80 and above. Age is the major risk factor for prostate cancer mortality and the proportion of deaths from prostate cancer in men aged 80 years plus is rising. In 2001, men aged over 80 years accounted for 48% of prostate cancer deaths, but by 2017 this had risen to 60%[15].

When compared to other cancers with screening programmes the age-standardised mortality rates are higher than breast, colorectal and cervical cancer, which have established screening programmes. The mortality rate for prostate cancer has also remained high while other cancers have seen a consistent decrease in mortality rates over the long term. Between 1971 and 2017 there was a 20% increase in the age standardised mortality rate of prostate

cancer. Although this has been gradually decreasing over the last decade it has not yet returned to levels prior to 1980s. Figure 2 compares the relative mortality rates for prostate, lung, breast, colorectal and cervical cancers. Prostate cancer remains the anomaly compared to these common cancers that have screening programmes given the overall increase in mortality rate.

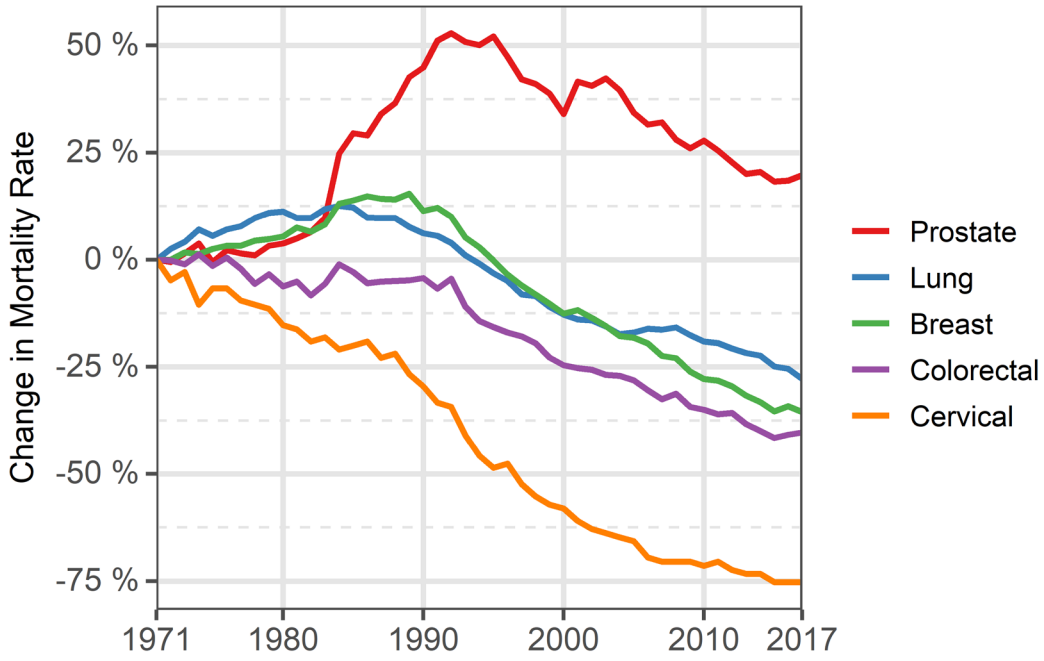


Figure 2: Relative change in Age-Standardised Mortality Rates for major cancers , 1971 to 2017 [11]. Relative to age-standardised mortality rate at baseline in 1971. Mortality rates per 100,000 men/women, standardised using the 2013 European Standard Population. Malignant neoplasm of the prostate coded as C61 from the International Classification of Diseases, Tenth Revision (ICD-10). Malignant neoplasm of the lung and trachea (C33-C34), breast (C50), colon and rectum (C18-C20) and cervix uteri (C53).

Due to the problems of establishing causation from observational data, the reasons for the changing trend in age-specific prostate cancer mortality has been widely debated. The reduction in age-standardised mortality rates over the last decade has often been attributed to the development of PSA testing. However, similar reductions in age-standardised mortality rates occurred across most developed countries irrespective of the extent of PSA screening[13]. Therefore, it is argued that this reduction is driven by improvement in treatments rather than prostate cancer screening[13]. It was during this period that radical prostatectomy was developed and androgen deprivation therapy in combination with radiotherapy became widespread.

In contrast, the rise in mortality rates in the mid-1980s has been argued as being caused in part by attribution bias from the discovery of PSA[16]. This attribution bias occurs if there is a misclassification of cause of death in men with low-grade (insignificant) prostate cancer who

die from other causes but who mistakenly are labelled with prostate cancer as a contributory cause during death certification. Nevertheless, the high rates of prostate cancer mortality clearly warrant consideration as a disease suitable for screening. It is likely that this will become more relevant over the coming years due to projected increases in morbidity and mortality[6].

2.4 Natural history of 'insignificant' prostate cancer

"The natural history of the condition, including development from latent to declared disease, should be adequately understood"
World Health Organisation [10]

Screening aims to reduce mortality rates through early detection of disease when it is more amenable to treatment. This requires an understanding of the natural history of the condition to identify disease at an early stage. Diseases with long preclinical periods are more likely to benefit from screening than those with short preclinical periods. In contrast, a screening programme will have minimal benefits if it detects cancer at a stage when it is incurable.

In principle, this should be an advantage in prostate cancer given the slow progression of the disease in the majority of cases. The dilemma for prostate cancer stems from the high prevalence of subclinical prostate cancer in the screening population. Autopsy studies have shown a substantial age-related prevalence of latent prostate cancer[17]. The potential for screening to inadvertently detect this latent disease is high and the risk of overdiagnosis, and subsequent over-treatment that leads to harm, remain the key concerns preventing recommendations for population-based screening[18]. It has been argued that these low-grade and low-volume lesions do not have the typical hallmarks of cancer and certainly do not behave aggressively and may be regarded as clinically insignificant[19].

There remains a lack of consensus on the definition of latent, or 'insignificant', prostate cancer. The most widely used definition of clinically insignificant disease was based on the histopathological parameters set out by Stamey[20] and Epstein[3]. Insignificant prostate cancer was defined on whole-mount prostatectomy as an organ-confined tumour with volumes ranging from 0.2cm^3 to 0.5cm^3 and no Gleason patterns 4 or 5. The original paper by Stamey et al[20] described a single parameter of tumour volume ≥0.5cm^3 from a cystoprostatectomy series based on an 8% lifetime risk of being diagnosed with clinically significant cancer. Epstein et al reported a volume threshold of <0.2cm^3 as being insignificant if the criteria of no capsular penetration was applied.

However, this definition has been generally considered too stringent and a new definition of <0.5cm^3 has been applied as the threshold for insignificant disease. In recent years, there has been a growing consensus that the 0.5cm^3 volume threshold also remains too conservative. In a contemporary cystoprostectomy cohort applying the Stamey criteria, Winkler et al[21] identified a higher threshold of 1.09cm^3. An analysis of the radical prostatectomy specimens from the European Randomized Study of Screening for Prostate Cancer found that a volume up to at least 1.3cm^3 for Gleason Score 6 could be included in the definition of insignificant disease[22].

In recent years, the independent prognostic value of tumour volume has been called into question. The majority of experts agree that Gleason score and pathological stage are more significant prognostic variables than tumour volume[23]. There have only been a few studies which have shown that tumour volume provides prognostic information independent of Gleason score, pathological stage and surgical margin status[24-26]. Although tumour volume correlates with these features at radical prostatectomy, when considered in a multivariate analysis model the majority of studies have not been able to demonstrate that tumour volume is an independent prognostic factor[27-31]. There is also no accepted standard for measurement of tumour volume; a range of approaches from 3D volume reconstruction to naked eye examination and a subjective descriptor such as small or large volume tumour.

2.5 Risk stratification

"There should be a recognizable latent or early symptomatic stage."
World Health Organisation [10]

There is a clear need for a reliable methods of risk stratifying men so that the correct disease can be identified at an early stage and treatment be directed towards those who are more likely to derive a cancer-specific mortality benefit. At present, tumour grade at biopsy has been the most widely utilised prognostic factor[32] in combination with other parameters such as PSA, clinical stage and tumour volume[33, 34].

The evidence for which disease will benefit from active treatment is evolving. As previously explained there is consensus that the early definitions of significant disease were too relaxed by including low volume Gleason 3+3 (ISUP Grade Group 1). Even this threshold may be too conservative and low-volume intermediate risk disease has increasingly been considered suitable for active surveillance in national guidelines[35, 36].

This was supported by the 10 year outcomes of the Prostate Testing for Cancer and Treatment (ProtecT) trial which reported minimal prostate cancer mortality in the active monitoring arm despite 22% having Gleason ≥7 disease[37]. The recent 29 year update of the Scandinavian Prostate Cancer Group Study 4 (SPCG-4)[38] which randomised between radical prostatectomy or watchful waiting showed that the risk of death from prostate cancer was similar between Gleason 3+3 and Gleason 3+4, with only the presence of Gleason ≥4+3 independently predicting prostate cancer specific mortality risk.

Given that a screening test for prostate cancer should be calibrated to detect the subset of cancer which will cause morbidity or mortality if left undetected and untreated, it is important to establish the target disease for screening. A variety of different definitions for clinically significant prostate cancer have been proposed using different clinical parameters[34, 39].

Table 1: Definitions for significant cancer

Definition	Parameters
Epstein criteria*[39], 1994	- Tumour volume ≥0.5 cm^3 or - Presence of Gleason ≥4 or - Non-organ confined disease
ISUP[40], 2005	- Any Gleason ≥ 3+4 (ISUP ≥2)
UCL/Ahmed 1[41], 2011	- Gleason ≥ 4 + 3 (ISUP ≥3) and/or - Maximum cancer core length (MCCL) ≥ 6mm
UCL/Ahmed 2[41], 2011	- Gleason ≥ 3 + 4 (ISUP ≥3) and/or - Maximum cancer core length (MCCL) ≥ 4mm

Abbreviations : ISUP = International Society of Urological Pathology. * Epstein criteria was originally designed for definitions of insignificant disease. Adapted here for significant disease

The Epstein criteria has been largely replaced with a more general definition of any Gleason ≥ 3+4 (ISUP ≥2) and many studies evaluating new screening tests have adopted this threshold [42, 43]. However even this threshold may be too conservative and low-volume intermediate risk disease has been considered suitable for active surveillance in national guidelines[35, 36]. In addition other factors such as tumour volume are not included in the standard definitions which is controversial as traditionally volume-based parameters have been highly predictive of disease progression[44]. There are alternative definitions such as UCL/Ahmed 1 & 2 which incorporate a combination of Gleason Grade and tumour volume. As there is no consensus on the definition for the purposes of this dissertation, a range of definitions will be reported.

2.6 Prostate Specific Antigen (PSA)

"There should be a suitable test or examination"
World Health Organisation [10]

2.6.1 History of PSA

PSA was originally developed for prostate cancer surveillance in the 1980s[45]. At the time the digital rectal examination (DRE) was the only test available to diagnose early prostate cancer. This test has high inter-examiner variability and limited sensitivity for early prostate cancer[46]. It does not allow examination of the whole prostate and many cancers detected by DRE are already locally advanced[47]. The widespread adoption of PSA screening began when studies showed that PSA had a better diagnostic performance for early prostate cancer than DRE[48].

However, its performance compared to an inadequate test does not necessarily translate into an ideal screening test. Before a screening test can be introduced the potential benefits need to be shown to outweigh the harms. The key measure of a screening tests' benefit is a reduction in disease specific mortality. Despite no evidence that PSA screening improved prostate cancer mortality, by 2001 a population-based survey from the US found that 75% of men over 50 years were having a PSA test[49].

Given the rapid adoption of PSA prior to the evidence, it is not surprising that the last two decades has witnessed a long and rigorous debate regarding the value of mass population PSA screening. This debate has included recurrent discussion on the optimal threshold to denote a screen-positive PSA test. The early studies utilised a threshold of 4ng/ml but subsequent work suggested that this was associated with an inadequate sensitivity and the threshold was progressively lowered to 3ng/ml during the 1990s.

The reduction in the threshold and widespread adoption of PSA screening has driven a rapid growth in the incidence of prostate cancer. Since the introduction of PSA the incidence of prostate cancer has doubled in the UK. A distinct geographical variation has emerged in countries where the test is accessible with incidence rates fluctuating up to 50-fold worldwide[50]. Yet the mortality rates between countries show less variation[51] and a substantial proportion of the rising incidence has been attributed to overdiagnosis caused by PSA screening.

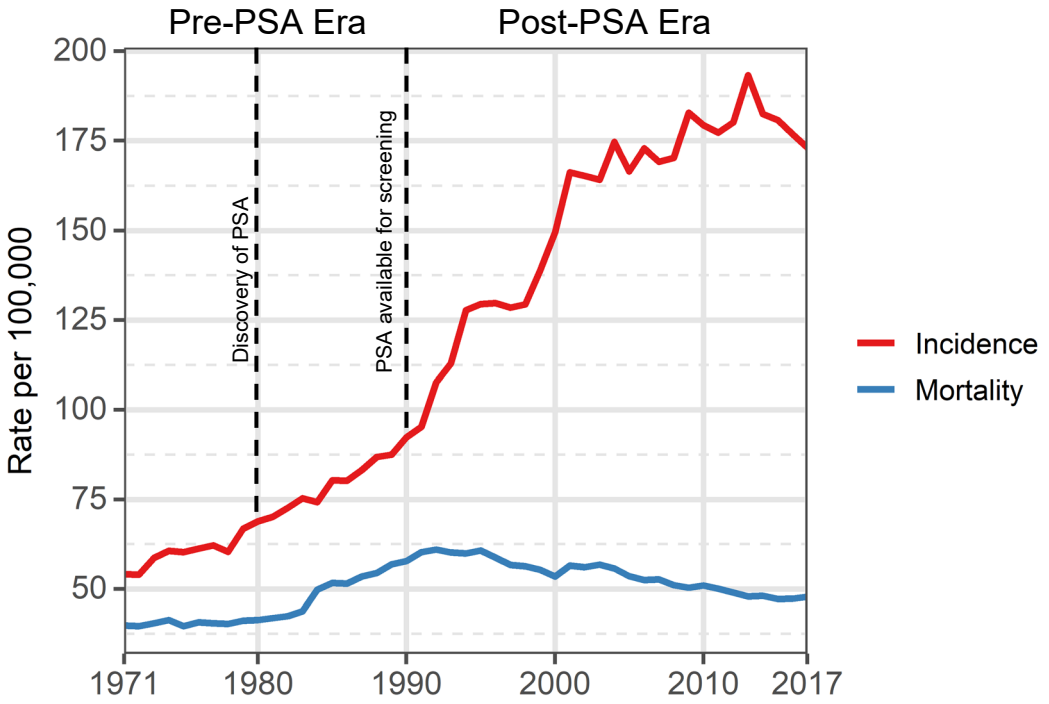


Figure 3: Prostate Cancer Incidence and Mortality Trends, 1971 to 2017. Age-standardised rates per 100,000 men, standardised using the 2013 European Standard Population. Cancer incidence rates are based on new cases registered in each calendar year. Source: Office of National Statistics[52]. Mortality includes deaths where prostate cancer was the underlying cause of death. Source: CRUK[11].

In certain diseases, incidence can be reduced by primary prevention such as changes in lifestyle or medication. This is occurring in cervical cancer with the introduction of human papillomavirus vaccine and in lung cancer with smoking cessation. There has been limited success with primary prevention in prostate cancer as the major risk factors are non-modifiable. Older age, family history and ethnicity remain the only well-established risk factors although there is some evidence for dietary, body size and androgen-related factors as weak risk factors for prostate cancer[53].

Attempts at primary prevention using finasteride have had limited success. The landmark Prostate Cancer Prevention Trial (PCPT) evaluated whether finasteride could reduce the risk of prostate cancer given that the disease is known to be androgen dependent[54]. The results paradoxically showed that the risk of high-grade cancer was increased with finasteride although the relative risk of all prostate cancer was 24.8% lower than with placebo. This has subsequently been shown to be an artefactual problem with biopsy of smaller prostates finding it easier to hit the higher grade cancers present.

2.6.2 Effectiveness

There have been six randomised controlled trials on PSA screening and each has considerable differences with respect to the trial settings, population, screening regimens, PSA thresholds and follow-up (Table 2). Many of the studies have been judged by a Cochrane systematic review to have methodological limitations[55] and in combination have failed to provide consistent results on the effects on prostate cancer mortality.

The Cochrane meta-analysis found only one trial to be at lower risk of bias, the European Randomized Study of Screening for Prostate Cancer (ERSPC). The ERSPC study has reported a 20% relative reduction in prostate-cancer mortality at 16 years[56] which corresponds to one prostate cancer death averted for every 570 men invited to screening. The numbers needed to screen have improved with longer follow up but they have not yet reached a similar level to mammography for breast cancer screening[57].

The ERSPC study was conducted across eight European countries with diverse inclusion criteria, screening protocols and randomisation approaches. The benefit from screening was not consistent between countries and significant mortality reductions were only demonstrated in Sweden and the Netherlands. The heterogeneity in outcomes may be influenced by variations in screening intensity[58] and the largest reduction in prostate cancer mortality was at the Swedish site with the most intensive biennial screening schedule. The remaining countries did not show a significant mortality benefit. Nevertheless, these findings have been highlighted as evidence that certain approaches to prostate cancer screening can have a mortality benefit although there may be a high cost of false-positive findings, overdiagnosis and overtreatment.

It is the mortality benefits from the ERSPC study which form the justification for international guidelines concluding that PSA screening may yield a small reduction in prostate cancer[18, 59-61]. There is a consensus that other trials should not be used as a primary source to draw conclusions for or against PSA screening due to their methodological limitations. Notably, the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial reported no mortality benefit with up to 19 years follow-up[62] but had a high level of PSA pre-screening, major control group contamination and low biopsy adherence in the intervention arm[62].

The PLCO trial was performed in the US between 1993 to 2001 during an era when PSA testing was becoming a standard test for most men over 50 years. Prior to enrolment 44% of participants had a previous PSA test and during the study 86% of the non-screened arm

underwent some PSA screening and more than 50% had annual screening[62]. This crossover has limited the power of the PLCO study to detect any significant difference in prostate cancer mortality[63] and it is not expected that a mortality benefit will emerge even with further follow-up.

In contrast, the ERSPC study benefited from being completed during a period when background PSA testing was low in Europe and the contamination has been estimated at 15%[64]. Micro-simulation studies have attempted to reconcile the differences between the ERSPC and PLCO. After adjusting for major confounders such as screening protocols, participant adherence, contamination and screening intensity, the model suggests that both studies could be compatible with a 25-30% lower risk of prostate cancer mortality[65]. Other studies have applied the same PLCO contamination and compliance rates to ERSPC and found that the mortality benefit in ERSPC substantially reduces to between 6% to 8% which is much closer to rates in PLCO[66].

More recent RCTs such as the UK Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP)[67] and a trial in Stockholm[68] evaluated the effect of a low-intensity, once only PSA test and found no effect on mortality. It is likely that an effective PSA screening programme requires repeat testing at short intervals with a trade-off of greater harm from false positives and overdiagnosis[69].

Table 2: Overview of randomised screening trials

	CAP[67]	ERSPC[56] (core)	Norrköping[70, 71]	PLCO[62]	Quebec[72]	Stockholm[68]
Start date	2001	1993	1987	1993	1988	1988
Study area	Primary care, UK	8 European countries	Single city, Sweden	10 centres, USA	Single city, Canada	Single city, Sweden
Randomisation method	Cluster	Individual	Individual (Quasi random)	Individual	Individual	Individual
Population						
Age range (years)	50 - 69	55 - 69	50 - 69	55 - 74	45 - 80	55 - 70
Source	Primary care	Population registry	Population registry	Volunteers	Population registry	Population registry
Screening Arm	189 386	72 890	1 494	38 340	7 348	2 400
Control Arm	219 439	89 351	7 532	38 343	14 231	25 081
Screening protocol						
Screening tests	PSA	PSA ± DRE	DRE ± PSA*	PSA + DRE	PSA + DRE ± TRUS	PSA + DRE +TRUS
PSA threshold	≥3ng/ml	2.5ng/ml to 4ng/ml**	≥4ng/ml	≥4ng/ml	≥3ng/ml	≥10ng/ml
Screening Interval	Single screen	2 years to 7 years***	3 years	1 year	1 year	Single screen
Outcomes						
Follow-up (years)	10	16	20	15	11	20
PCa mortality per arm (Screened / control)	549 / 647	520 / 793	30 / 130	255 / 244	10 / 74	86 / 771
Rate ratio (95% CI, p value)	0.96† (0.85-1.08 p = 0.35)	0.080† (0.72-0.89 P < 0.001)	1.23‡ (0.94-1.62 p = 0.13)	1.04† (0.87-1.24 p = 0.67)	1.09‡ (0.82-1.43 p = 0.56)	1.05‡ (0.83-1.27 p = NR)
Risk of bias****	Higher	Lower	Higher	Higher	Higher	Higher

Abbreviations: NR = Not reported, PCa = Prostate cancer, TRUS = Transrectal ultrasound.

* PSA testing commenced in 1993. Screening rounds 1-2 were with DRE only.

** The standard threshold was 3ng/ml. In certain countries, ancillary tests were used for PSA 2.5-3.9ng/ml.

*** The screening interval was 4 years in most countries except Sweden (2 years) and Belgium (4 to 7 years).

**** As judged by a Cochrane systematic review[55], updated in BMJ (2018)[73].

† Rate by total number of person-years, ‡ Rate by total number of participants.

2.6.3 An uncertain benefit-harm ratio

While the evidence that PSA screening reduces prostate cancer mortality is variable, the harms from PSA screening are clear. The most serious harm associated with PSA-based screening programmes is overdiagnosis. The estimates of overdiagnosis vary from 1.7% to 67%[74] although in the ERSPC study it was estimated that 50% of cancers were over-diagnosed and would not have caused any morbidity or mortality during the man's lifetime[75]. As the incidence of insignificant cancer is markedly age-related and Western countries have an aging population, it is expected that the rates of insignificant cancer will dramatically rise[76]. In large autopsy studies the risk of incidental prostate cancer increases in a non-linear fashion, from 15% of men aged 40-50 years to 59% by age 79 years or more.

The consequences of overdiagnosis are a major problem as men are subjected to the psychological distress of an unnecessary cancer diagnosis and to the morbidity from the consequent overtreatment. This overtreatment is also associated with additional healthcare cost and side effects including long-term risks from sexual, urinary and bowel complications[77]. The adverse event pattern varies depending on the type of treatment but the patient-reported outcomes in the ProtecT trial have highlighted that there are consistent issues with urinary incontinence, sexual dysfunction and bowel problems[77]. Microsimulation models within ERSPC have demonstrated that although there may be a 20% relative reduction in prostate-cancer mortality, this benefit is offset by a 23% reduction in life years conferred by treatment-related side effects[78].

The problem is compounded because the benefits from active treatment of PSA-screen detected cancers remain uncertain. The Prostate Testing for Cancer and Treatment (ProtecT) study showed that prostate cancer-specific mortality was very low (1.5 deaths per 1,000 person-years) for screen-detected prostate cancer after 10 years follow-up[37]. There was no significant difference in mortality between the treatment and active monitoring arms although the risk of metastasis was higher in the active monitoring arm. The PIVOT study confirmed this for men diagnosed early in the PSA screening era in the USA but showed a potential benefit in the intermediate and high-risk subgroups [79].

Due to these findings, active surveillance has become an important option to reduce harms from overtreatment. However, uptake rates are highly variable and contemporary registry data has shown that acceptance of active surveillance for low-risk prostate cancer is 96% in UK[80], 74% in Sweden[81], 66% in Australia[82], 57% in Canada[83] and 40% in USA[84]. Although the

long-term outcomes from active surveillance are encouraging, a significant proportion of men still exit active surveillance despite no evidence of disease progression and suffer the negative effects of overtreatment[85]. Active surveillance as a strategy can also confer direct healthcare burden and harms through use of repeated biopsy every 1-2 years.

2.7 Defining a screen-positive PSA

"There should be an agreed policy on whom to treat as patients"
World Health Organisation [10]

The Wilson and Jungner criteria include a further principle that a suitable cut-off level has to be set within a screening programme to identify patients who require further investigations. This has been re-emphasised by the UK national screening committee which set out guidelines for appraising the appropriateness of a screening programme; they stated that "a suitable cut-off level (must be) defined and agreed".

A suitable cut-off is the one which yields an appropriate balance between sensitivity and specificity. There have been many studies reporting variable sensitivity-specificity profiles for PSA; however, most suffer from verification bias as the status of men with a low PSA are not confirmed with biopsy. The Prostate Cancer Prevention Trial (PCPT)[86] is a unique large-scale study as all participants underwent biopsy irrespective of PSA. It provides a valuable insight into the true diagnostic accuracy in a screened population. At the standard PSA threshold of 3ng/ml, the sensitivity of PSA for clinically significant prostate cancer was only 57.6%.

These findings have been confirmed in other studies which utilise mathematical modelling to correct for verification bias[87]. Given that the reference standard in these studies was transrectal ultrasound guided (TRUS) biopsy, which is known to have a significant sampling error and to underestimate disease burden, with a reported sensitivity of 48% itself, the true sensitivity of PSA-TRUS biopsy pathway is likely to be even lower[88, 89].

Due to the modest sensitivity of PSA, international guidelines have been unable to provide a threshold recommendation as there is a risk of missing significant cancer at all levels of PSA[90]. Indeed, in the Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) there were 146 men who died from prostate cancer despite attending the one-off PSA screening clinic. Of those who died from prostate cancer 68 (46.6%) had a PSA level ≤ 3ng/ml on screening.

An improvement in the sensitivity of PSA could be achieved by lowering the screening threshold. However, the PCPT trial showed that to reach a sensitivity similar to

mammography, a PSA threshold of 1.6ng/ml would be required. At this threshold, the trade-off is a low specificity of 54.8% which would lead to an increase in biopsy referrals, higher false positives rates, an increase in biopsies, and a rise in overdiagnosis and over-treatment of insignificant disease[91]. This would occur if the reflex next step in the pathway after a 'positive' PSA is a biopsy. Later in my thesis I will discuss the changes occurring in secondary care which are likely to impact on this.

Regardless, the appropriate PSA threshold for screening remains controversial and there is no level which yields an optimal balance between sensitivity and specificity. The test will result in either high false-positives or false-negatives depending on the cut-off value. There is continuing debate on the optimal threshold for screening, but it cannot overcome the fundamental problem that PSA is not a dichotomous test and is more useful as a continuous variable reflecting a continuum of prostate cancer risk. Indeed, when the validity standards of PSA were assessed within a prospective nested case-control study, there was no PSA threshold which attained the likelihood ratios required of a screening test[92].

Table 3: Criteria for appraising the viability, effectiveness and appropriateness of a screening programme as per UK National Screening Committee

The condition

1. The condition should be an important health problem.
2. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease.
3. All of the cost-effective primary prevention interventions should have been implemented as far as practicable.

The Test

1. There should be a simple, safe, precise and validated screening test.
2. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.
3. The test, from sample collection to delivery of results, should be acceptable to the target population.
4. There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.

The Treatment

1. There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
2. There should be agreed evidence-based policies covering which individuals should be offered treatment and the appropriate treatment to be offered.

The screening programme

1. There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity.
2. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.
3. The benefit gained by individuals from the screening programme should outweigh any harms, for example from overdiagnosis, overtreatment, false positives, false reassurance, uncertain findings and complications.
4. The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criteria should have regard to evidence from cost benefit and/or cost effectiveness analyses and have regard to the effective use of available resources.

2.8 Acceptability of PSA

"The test should be acceptable to the population"
World Health Organisation [10]

The value of PSA as a screening test lies in its simplicity, reproducibility, lack of invasiveness and low cost. However a high rate of false positives is a known issue in PSA-based screening programmes due to the prevalence of other PSA-elevating prostate diseases, particularly benign prostate hyperplasia (BPH) and (non-infective) prostatitis, within the screened population[93]. This is more likely if the PSA is between 0-10ng/ml as the specificity at this level is limited. Studies of cystoprostatectomy specimens evaluating the prevalence of prostate cancer have suggested that the majority of PSA up to 10ng/ml is produced by BPH rather than prostate cancer[21]. Longitudinal follow-up within the ERSPC study has shown that the false positive rate is 17.8% over three screening cycles[69].

Excessive false positives cause significant downstream harms by triggering a large number of unnecessary prostate biopsies. In the Göteborg arm of the ERSPC study, the intensive PSA screening protocol led to nearly half of men receiving a false positive biopsy recommendation over six screening rounds[69]. These unnecessary biopsies increase the costs of the screening programme and expose men to biopsy-related morbidity.

Transrectal ultrasound-guided (TRUS) biopsy remains the standard biopsy technique within screening programmes although the use of transperineal biopsy is increasing in the UK. TRUS is associated with serious morbidity from infectious complications. Indeed, the incidence of hospitalisation from serious infectious complications is rising due to higher rates of fluoroquinolone-resistant organisms[94]. Other side effects reported in ERSPC were haematospermia (50.4%), haematuria (22.6%), pain after biopsy (7.5%), fever (3.5%), and urinary retention (0.4%)[95].

In addition, the majority of guidelines for PSA-screening requires PSA to be combined with a digital rectal examination (DRE) during the initial screening[48]. DRE has been identified as a key barrier to attending prostate cancer screening due to its invasive nature. A study comparing the willingness of men to participate in a PSA screening in combination with DRE found that DRE prevented 22% of men from participating in prostate cancer screening[96].

2.9 Current status of PSA screening

"Case finding should be a continuous process."
World Health Organisation [10]

The uncertainty regarding the benefits and harms of PSA screening has led to most Western countries recommending against routine PSA screening in favour of opportunistic screening (Table 4). National guidelines emphasise that opportunistic PSA screening should only be performed following a shared decision-making process[35]. This requires patients to reach a tailored decision based on their own values after detailed discussion to understand the trade-off between benefits and risks.

In clinical practice it has been challenging to deliver such an individualised process across a large population[97]. Decision aids have been designed to improve information quality, decrease physician bias and reduce consultation time[98]. A meta-analysis of seven PSA decision-aids found that men were less likely to have a PSA test after receiving information on the benefit and harms of PSA screening via a decision aid[99].

The effectiveness of a screening programme is dependent on uptake and participation in follow-up investigations. The uptake of opportunistic PSA screening is highly variable between countries[100] and there are further disparities among different population subgroups[101]. In the UK, the penetration of PSA screening in the population is approximately 6% per annum[101, 102] and participation rates are low among ethnic minorities and in deprived communities[101]. In Europe, the uptake of patient-initiated PSA screening is 4% in men in the lowest socioeconomic category compared to 10% in the highest[100]. Similar disparities in socio-economic participation are not found in population-based screening programmes for other cancers across Europe[103]. The low and unequal uptake of opportunistic PSA screening makes it unlikely that this method of screening will have any significant impact on population level mortality rates.

Table 4: Screening guidelines from major organisation in the USA and Worldwide	
Screening Test	**Recommendation**
United States	
US Preventative Task Force (USPSTF)[104]	1. Informed men 55-69 years about potential benefits and harms so they can make an informed, personal decision about whether to have PSA screening 2. Do not screen men age ≥ 70 years
American Cancer Society (ACS)[105]	1. Men age >50 with > 10-yr life expectancy should be given an opportunity to make an informed decision about whether to have PSA screening 2. Re-screening interval depends on level of baseline PSA
National Comprehensive Cancer Network (NCCN)[106]	1. After appropriate counselling offer PSA to men 45-75 years 2. Re-screening interval depends on PSA baseline
American Urological Association[107]	1. Routine PSA testing not recommended 2. Shared decision-making for men age 55 to 69 years considering a PSA test
Worldwide	
European Association of Urology (EAU)[108]	1. Does not recommend screening in men 40-69 years at average risk 2. Offer an individualised risk-adapted strategy to a well-informed man with a life expectancy of at least 10–15 yr.
UK National Screening Committee (UKNSC)[109]	1. Recommends against national PSA screening 2. Informed decision-making process for men over 50 years who request a PSA test
New Zealand Prostate Cancer Taskforce (PCT)[110]	1. Does not recommend a national screening programme 2. informed decision process with men aged 50 to 70 years
Prostate Cancer Foundation of Australia (PCFA)[111]	1. Does not recommend a population screening programme 2. Offer decisional support to men age 50-69 and a PSA test after discussion of benefits and harms
Canadian Task Force on Preventive Health Care[112]	1. Recommends not screening for prostate cancer with the PSA test, applies to men of all ages

2.10 Treatment for significant prostate cancer

"There should be an accepted treatment for patients with recognized disease"
World Health Organisation [10]

Theoretically an screening programme might allow clinically significant prostate cancer to be identified when it is more amenable to certain minimally invasive treatments. The most established minimally invasive treatment for prostate cancer is known as focal therapy which is designed for men with localised, low volume intermediate risk disease.

Focal therapy encompasses a wide range of approaches that allow selective ablation of target areas. This may be delivered by a variety of energy modalities including high-intensity focused ultrasound (HIFU), cryotherapy, photodynamic therapy, focal laser ablation, focal brachytherapy, irreversible electroporation and radiofrequency ablation as well as interstitial drug injections.

The aim of focal therapy is to retain equivalent oncological outcomes to whole-gland therapies while reducing the side-effects associated with these treatments. In the absence of screening, prostate cancer often presents at a later stage with a disease pattern which is less amenable to minimally invasive treatments. This was seen in the PIVOT trial which was conducted in the pre-PSA era and 12% had high-risk disease so would not be suitable for focal therapy. In contrast, the ProTect trial was conducted using PSA as screening test and found a higher proportion of participants with intermediate risk disease.

A potential benefit of prostate cancer screening could be to increase the number of cancers identified suitable for focal therapy and potentially reduce the morbidity associated with other treatments. The functional outcomes of focal therapy have been summarised in a meta-analysis comparing patient reported outcomes measures of different whole gland therapies with HIFU[113]. This showed that there was no significant deterioration of sexual function or incontinence at one year. Similar functional outcomes were found by Yap et al who reported that although potency deteriorated at one and three months post HIFU it returned to baseline by six months[114].

For high-risk prostate cancer, there are alternative whole-gland therapies available such as surgery and radiotherapy. The PIVOT and SPCG-4 RCTs showed that the benefit of treatment occurs in either high-risk or intermediate-risk men[38, 79]. In PIVOT, men with intermediate risk disease who underwent radical prostatectomy gained 2.1 years of life (95% CI 0.4-3.7) to the observation group after 22 years follow-up[79]. In SPCG-4, which included men with clinically

detected prostate cancer, a mean of 2.9 additional years were gained from radical prostatectomy.

While whole-gland treatments provide effective cancer control, they may cause more side effects as seen in SPCG-4 where men in the radical prostatectomy had greater erectile dysfunction compared to those assigned to watchful waiting[115]. In ProtecT, detailed quality of life data was collected for urinary, bowel and sexual function in additional to general quality of life measures[77]. Radical prostatectomy had the greatest adverse effect on urinary continence and sexual function.

2.11 Conclusion

Few issues have generated as much controversy as prostate cancer screening and this chapter has summarised some of the key challenges and critical issues. The current status of PSA-based screening is complex and the evidence from large randomised screening trials has been contradictory[56, 62]. There remains a lack of consensus on whether the harms of overdiagnosis and over-treatment outweigh the potential reductions in prostate cancer specific mortality. This debate on the role of PSA for screening has continued for more than two decades and is likely to continue unresolved until alternative screening modalities are developed which can differentiate the presence or absence of clinically significant prostate cancer at a population level.

Chapter 3 – Challenges and considerations for MRI as a screening test: A narrative review

3.1 Overview

The uncertainty around PSA as a screening test raises the question of whether an alternative screening modality could achieve a more acceptable balance between benefits and harms. As detailed in the background chapter, there is a particular need for a screening test which can reduce the burden of overdiagnosis and improve detection of clinically significant cancer.

I completed a comprehensive narrative review to identify key themes relevant to MRI and screening for prostate cancer. The aim was to synthesise a broad range of potential benefits and weigh these against potential harms from MRI screening. This chapter forms the basis of work published in Nature Reviews Urology[ii].

3.2 Introduction

The previous chapter highlighted how screening tests have to meet criteria originally outlined by Wilson and Jungner[10] and subsequently adopted and built upon by the UK National Screening Committee[116]. These include being simple to perform, having a high sensitivity and specificity for the disease, being acceptable to the target population and having an agreed threshold for triggering further diagnostic investigations. Whilst PSA screening does not meet some of these key principles, the onset of advances in imaging technology have allowed alternative modalities to be considered that might.

The use of image-based screening tests has expanded with the technological advances in imaging techniques and protocols. Mammography for breast cancer is an established screening programme across the majority of developed countries and there is mounting evidence for the efficacy of CT colonoscopy for bowel cancer screening[117]. Screening programmes of high risk individuals have been established using low-dose helical CT (LDCT) for lung cancer and MRI for breast cancer[118].

[ii] Eldred-Evans, D., et al. (2020). "Rethinking prostate cancer screening: could MRI be an alternative screening test?" Nature Reviews Urology **17**(9): 526-539.

Advances in imaging technology have resulted in the development of multiparametric MRI which combines anatomical T2-weighted imaging with functional techniques such as diffusion-weighted imaging (DWI), dynamic contrast-enhanced (DCE) sequences and/or MR proton spectroscopy. The clinical applications of MRI have expanded to include pre-biopsy triage, tumour detection, localisation, staging, surveillance and assessment of reoccurrence as well as image guidance for biopsy, focal therapy, radiotherapy and surgery.

However, its role as a screening test has not been widely considered and this narrative review explores the key literature relevant to MRI as a potential screening test. The purpose is to lay the foundations for the work presented in subsequent chapters and identify key themes related to MRI which need to be addressed in my thesis.

3.3 Methods

A systematic search of the literature was conducted in December 2017 to identify empirical studies of MRI-based screening for prostate cancer. The databases used included PubMed, Embase, and Cochrane. Each database was searched from inception up to December 21, 2017, with no language restrictions, using the terms “screening”, “prostate cancer” and “Magnetic Resonance Imaging”. The search resulted in 503 articles, with 301 remaining after duplicate removal.

An initial screening of the articles was conducted to identify those relevant to the topic. Inclusion criteria were a clinical trial evaluating MRI as an independent screening test. Studies were excluded if MRI was evaluated as a secondary test following a screening PSA. The review process yielded a single small pilot study of volunteers (n=47) who underwent a multi-parametric MRI as a screening test for prostate cancer[119]. In the two years since my original search, an updated search has not revealed any further studies meeting the search criteria.

Due to the paucity of literature on the topic of MRI as a screening test, a broader narrative review was conducted. This narrative review targeted the quantitative and qualitative literature on pre-biopsy MRI which would be relevant to a screening setting. Methodologically a narrative review is a useful synthesis method where there are “complex, multifaceted issues to be explored, understood and contextualised”[120]. A narrative review can be more wide-ranging in scope than systematic reviews and provides researchers with a methodological basis for synthesising evidence across a diverse literature.

The phases of this narrative review followed the steps described by Greenhalgh et al 2005[120]:

(1) Planning phase: The initial broad research question was agreed between the research team. The research question was "Could MRI be an alternative screening test to PSA?". Further details of the proposal and planning phase are provided in Appendix V.

(2) Search phase: A search was conducted to identify a broad range of papers relevant to MRI as a screening test. PubMed, Embase, and Cochrane databases were searched for English-language articles published in peer review journals from inception to February 2019. The papers were screened for relevance of MRI as a screening test. Publications were not assessed on quality grounds but any issues with quality were highlighted in the review.

(3) Mapping: Each primary study was evaluated for relevance to the research questions. Comparable studies were grouped together and reviewed with two questions in mind: (a) What are the potential benefits of MRI as a screening test? (b) What are the main problems with MRI for screening?

(4) Appraisal phase: The contents of each paper were reviewed focusing on material related to the advantages and disadvantages of MRI as a screening test. I identified and extracted potential benefits and risks raised in each paper that were relevant to the objectives of the review. Following data extraction, key themes were identified from the literature.

(5) Synthesis phase: A narrative account was made of each theme identified from the literature. There were 13 themes identified which included. (i) Improving sensitivity for clinically significant disease, (ii) Reducing the burden of overdiagnosis, (iii) Reducing the harms of underdiagnosis, (iv) Improving acceptability and uptake of screening, (v) Improving adverse events and compliance, (vi) A low specificity and high false positive rate, (vii) Risk of missing low-volume significant prostate cancer, (viii) Incidental findings, (ix) Reproducibility, (x) Safety, (xi) Acceptability, (xii) Cost Effectiveness, and (xiii) Capacity.

3.4 Results

3.4.1 Search

The broader search for the narrative review identified 1,879 articles of which 1056 were unique citations. From these, 908 were excluded at abstract review. Full-text screening was carried out on 148 articles of which 118 were excluded. In total, 30 systematic reviews or meta analyses were included which provided potentially relevant information about MRI as a screening test. The flow chart illustrating the search process is shown in Figure 4.

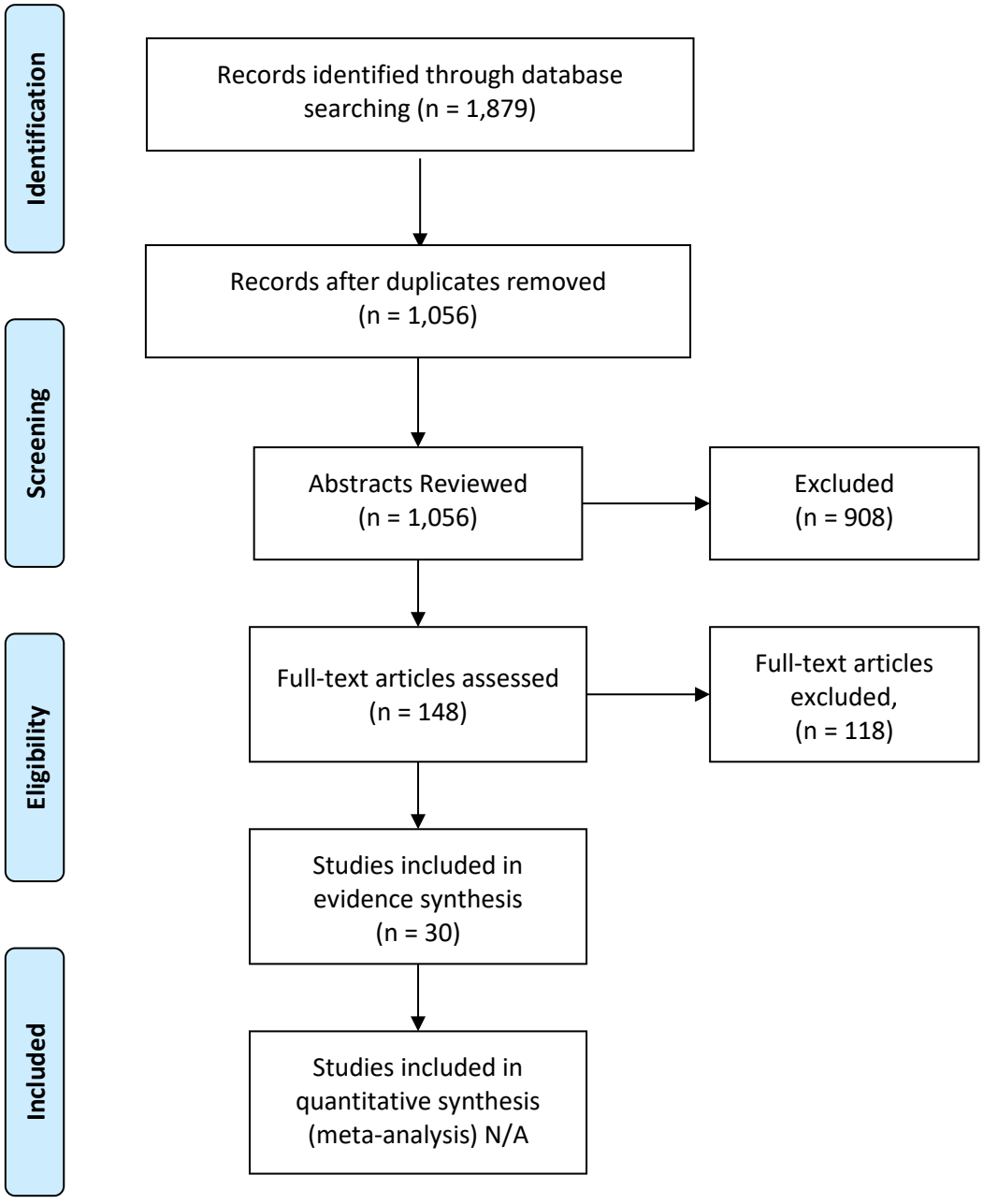


Figure 4: Flow diagram of included studies.

3.4.2 Evidence Synthesis

Table 5 shows summarises the 29 studies included in the narrative review. The studies were conducted between 2013 and 2019. The studies consisted of 5 meta-analysis, 3 systematic reviews, 3 randomised controlled trials of pre-biopsy MRI, 8 paired diagnostic accuracy studies, 8 prospective cohort studies, 1 discrete choice experiment and 1 cost-effectiveness modelling study.

Table 5: Characteristics of primary papers selected for narrative review

Author, Year	Design	Relevance / Findings
Ahmed, 2017[88]	Paired diagnostic study	Diagnostic Accuracy (secondary care)
Alberts, 2017[121]	RCT Diagnostic Utility	Effect on Overdiagnosis
Alberts, 2018[122]	RCT Diagnostic Utility	Adjunct Screening MRI
Boesen, 2018[123]	Paired diagnostic study	Diagnostic Accuracy of bpMRI
Bekker-Grob, 2013[124]	Discrete choice experiment	Trade-off between PCa and biopsy
Chen, 2017[125]	Diagnostic Meta-analysis	Diagnostic Accuracy of bpMRI
de Rooij, 2014[126]	Cost-Effectiveness Modelling	MRI targeted biopsy benefits
De Visschere, 2016[127]	Prospective cohort	False negatives
Drost, 2019[128]	Diagnostic Meta-analysis	Diagnostic Accuracy (secondary care)
Fütterer, 2015[129]	Systematic Review	Diagnostic Accuracy (secondary care)
Hamoen, 2015[130]	Diagnostic Meta-analysis	Diagnostic Accuracy (secondary care)
Jambor, 2019[131]	Paired diagnostic study	Diagnostic Accuracy of bpMRI
Kang, 2018[132]	Diagnostic Meta-analysis	Diagnostic Accuracy of bpMRI
Kasivisvanathan, 2018[133]	RCT Diagnostic Utility	Clinical Utility MRI Targeted biopsy
Miah, 2019[134]	Prospective Cohort	Impact of non-target biopsy
Moore, 2013[135]	Systematic review	Reducing overdiagnosis
Nam, 2016[119]	Prospective cohort	Diagnostic accuracy Screening MRI
Neves, 2018[136]	Paired diagnostic study	Diagnostic Accuracy of bpMRI
Niu, 2018[137]	Systematic review	Diagnostic Accuracy of bpMRI
Rastinehad, 2005[138]	Paired diagnostic study	Diagnostic Accuracy (secondary care)
Rosenkrantz, 2016[139]	Prospective cohort	Capacity of MRI
Rouvière, 2019[140]	Paired diagnostic study	Impact of non-target biopsy
Schouten, 2017[141]	Prospective cohort	False negatives
Stavrinides, 2019[142]	Prospective cohort	False negatives
Venderink, 2018[143]	Prospective cohort	Optimal MRI threshold
Weiss, 2018[144]	Paired diagnostic study	Diagnostic Accuracy of bpMRI
Wolters, 2011[145]	Prospective cohort	False negatives
Woo, 2017[146]	Diagnostic Meta-analysis	Diagnostic Accuracy (secondary care)
Woo, 2018[147]	Paired diagnostic study	Safety of bpMRI

Abbreviation: Prostate Cancer (PCa), Randomised controlled trial (RCT)

This evidence synthesis was divided into two sections: the potential benefits of MRI screening and problems of MRI screening. These have been further split into subcategories through broad thematic analysis.

3.4.3 Factors for MRI screening

3.4.3.1 Sensitivity for clinically significant disease

MRI has been shown to have a high sensitivity for clinically significant prostate cancer in men referred to secondary care due to a clinical suspicion (often elevated PSA or abnormal rectal exam) as demonstrated in several large diagnostic accuracy studies[88, 148] and systematic reviews[146, 149]. A Cochrane meta-analysis showed a pooled sensitivity of 91% (95%CI 0.83-0.95) across 12 studies which were selected against a strict criteria of template mapping biopsy or template saturation biopsy as a reference standard[128]. Indeed, within PROMIS study, which compared the diagnostic accuracy of MRI and TRUS-biopsy against a template mapping biopsy that sampled the prostate every 5mm, the sensitivity of MRI reached 100% for Gleason ≥4+3[88].

The PROMIS study provided high level evidence for the accuracy of pre-biopsy MRI compared to TRUS-biopsy using template prostate mapping biopsy as an accurate reference standard. Pre-biopsy MRI had a high sensitivity of 93% for the primary definition of clinically significant disease and this was consistent across a range of other definitions, including for any Gleason 3+4 disease where the sensitivity was 88%. Subsequent clinical utility studies have shown that use of pre-biopsy MRI reduces the number of men biopsied and the number of men diagnosed with insignificant cancers, with at least maintaining detection rates of significant cancer or increasing it[133].

Ongoing refinements to MRI techniques and interpretation through structured validated reporting schema such as the Prostate Imaging Reporting and Data System (PI-RADS) have shown improved sensitivity from 88% to 95% without a corresponding impact on specificity[146]. Beyond detection of significant disease, MRI provides valuable data on tumour location, volume and stage. Clinical utility studies have shown that the MRI pathway leads to 28% fewer biopsies when used as a triage test[133], reduces overdiagnosis (often nearly halving the number of Gleason 3+3 cancers) with similar or higher detection of significant cancers compared to traditional systematic TRUS biopsy[128].

This high sensitivity of MRI has been established within populations with a high prevalence of significant prostate cancer. Traditionally, sensitivity and specificity were assumed to be fixed test characteristics. However, it is increasingly recognised that sensitivity and specificity may vary across different populations with different disease prevalence, the so-called spectrum effect[150]. Rather than inferring the sensitivity of MRI based on the performance in a diagnostic setting, it needs to be specifically investigated in the general population where disease prevalence is lower.

There has been one small pilot study of 47 volunteers recruited via newspaper advertisements which compared the performance of PSA and multi-parametric MRI (mpMRI) against a reference test of 12-core systematic TRUS +/- targeted biopsy[119]. The area under the ROC curves (AUROC) suggested that mpMRI may be more accurate than PSA for detection of any prostate cancer (AUC 0.81 vs. 0.67). On multivariate analysis, MRI score was the only predictor of Gleason ≥7 (OR 3.5) compared to PSA (OR 1.0). The low prevalence of significant disease in the general population requires a much larger sample size to reliably compare the diagnostic accuracy of both tests. As well as its size, this study was limited due to potential spectrum bias since 36% had an abnormal PSA (≥4.0ng/ml), much higher than expected based on the 7.9% equivalent figure from the US PLCO screening study [151] and 11% in UK CAP with PSA ≥3.0ng/ml[67].

3.4.3.2 Reducing the burden of overdiagnosis

Screening programmes often use a two-step process with an initial test with high sensitivity followed up by a highly specific test to confirm the diagnosis. If a similar two-step approach was applied to MRI screening then a suspicious MRI would trigger an MRI targeted biopsy. This biopsy would be concentrated on the visually abnormal prostate areas given that there is no other cancer screening programme which incorporates a biopsy technique that samples apparently healthy tissue.

It is this non-targeted sampling of the gland that contributes to the high rates of overdiagnosis of insignificant disease. MRI targeted biopsy has been evaluated during the fifth screening round of ERSPC Rotterdam and the overdiagnosis rate was 7% compared to 28% in the 12-core TRUS biopsy arm[122]. Further, the MRI-FIRST study evaluated the need for non-targeted biopsy in a paired diagnostic study across 16 centres and found that overdiagnosis of insignificant disease was 20% in non-targeted cores compared to 5.8% in targeted cores[152]. In MRI-FIRST, the additional diagnostic yield from non-targeted cores was

5.2% and in other multi-centre series the detection of significant prostate cancer exclusively in the non-targeted cores has been as low as 1%[134].

This compares favourably with colposcopic biopsy in cervical cancer screening where non-targeted sampling of the normal appearing cervix finds an additional 4.5% of high-grade cervical intraepithelial lesions (CIN2+)[153]. In breast cancer screening, an MRI-targeted biopsy of the mammographically normal breast has an additional diagnostic yield of 3% in women with newly diagnosed unilateral cancer[154, 155]. If a contralateral prophylactic mastectomy is performed the detection of mammographically invisible cancer can increase to 6%[156]. Therefore, while non-targeted biopsy does detect a small number of additional prostate cancers, the additional diagnostic yield is similar to that found in other tumour groups where targeted-only biopsy is the standard of care.

3.4.3.3 Reducing the harms of underdiagnosis

Men undergoing PSA screening have to be informed that the risk of a false negative result is estimated to be 5%[157] to 15%[158]. False negatives reduce the public's confidence in the screening test and can cause diagnostic delays from false reassurance. MRI may improve this false negative rate given the recent Cochrane meta-analysis reported the pooled negative predictive value (NPV) for Gleason ≥3+4 as between 86% to 97% across different disease prevalences[128]. When the prevalence was set at 10%, the false negative rate of MRI reduced to 9 per 1,000 tested[128]. The predictive value of any test depends on disease prevalence within the population. In the screening population the prevalence of clinically significant disease will be much lower than in secondary care and in the PCPT trial was estimated as less than 5%[158]. Given the inverse relationship between negative predictive value and prevalence, it could be hypothesised that a screening MRI will have a sufficiently high NPV to safely exclude clinically significant disease and minimise unnecessary biopsy.

3.4.3.4 Improving acceptability and uptake of screening

On a population level, a screening test requires a high level of acceptability and uptake to deliver mortality benefits. MRI has characteristics which could improve participation rates and reduce inequalities in screening uptake. It is not affected by radiation risks which occur in mammography or low dose CT (LDCT) for lung cancer and it is less invasive than the current screening pathway which requires a digital rectal examination (DRE) to be performed in parallel with a PSA test[48]. A study comparing the willingness of men to participate in PSA screening in combination with DRE found that DRE would prevent 22% of men from

participating in prostate cancer screening[96]. The embarrassment and discomfort associated with DRE has been a key barrier to the uptake of prostate cancer screening among certain social and ethnic groups[159].

It is important that any screening test for prostate cancer is acceptable to high risk groups particularly black men who have the highest mortality rates from prostate cancer[160]. The effectiveness of PSA screening in black men is unknown as this group has been hugely under-represented in screening trials. There have been similar barriers to uptake described within flexible sigmoidoscopy screening programmes where the rectal nature of the procedure has been regarded as a threat to masculinity by African-Caribbean men[161]. In colorectal cancer screening, studies evaluating less invasive tests, such as CT or MRI colonography, have suggested that a non-invasive imaging test could improve acceptability and screening attendance[162, 163].

MRI screening programme would increase the uptake of targeted biopsy as abnormalities are visualised rather than seen on a serum biomarker that is tissue or organ specific but cannot localise cancer geographically to a zone of an organ. Targeted biopsy has been shown to provide better accuracy and characterisation of clinically significant disease, improved sampling efficiency and reduced histopathological burden[164]. A biopsy protocol which targets only the MRI lesion without systematic biopsy would reduce the risk of detecting insignificant cancer, which is highly prevalent within the majority of glands being sampled[135]. MRI-TRUS image fusion targeted prostate biopsies have been incorporated into the screening protocol in ERSPC Rotterdam resulting in a 50% reduction in overdiagnosis of insignificant disease[121].

3.4.3.5 Improving adverse events and compliance

The effectiveness of PSA screening was also limited by a high rate of non-compliance with TRUS-biopsy amongst men with a positive PSA test. In the PLCO trial, the proportion of men attending for a sextant TRUS-biopsy was consistently less than 40% across all four screening rounds[165]. In the ERSPC trial, compliance with biopsy recommendation was higher but variable between centres ranging from 39.8% to 94.2% within the baseline screening round[166]. In subsequent rounds of ERSPC, more than one third of men did not undergo the recommended biopsy and the overall rate of non-compliance was 14.4%.

This level of non-compliance makes it difficult to justify a screening programme based on TRUS biopsy. Although the reasons for non-compliance are not directly explored, it is likely

that factors such as poor diagnostic performance and biopsy-related morbidity are important, as shown in the UK CAP study[67].

A potential advantage of a screening programme using MRI-targeted biopsy is a reduction in adverse events and improved compliance. There is some evidence that the reduction in the number of cores in targeted biopsy protocols leads to less pain, fewer urinary symptoms and potentially lower rates of serious infection and sepsis, especially if a transperineal route is used[167]. Indeed, a discrete choice experiment of 1,000 men showed that they were willing to trade-off 2.0% or 1.8% risk reduction of PC-related death to decrease their risk of unnecessary treatment or biopsy by 10%, respectively[124].

3.4.4 Factors against MRI screening

3.4.4.1 A low specificity and high false positive rate

MRI has a relatively low specificity and in the Cochrane meta-analysis the pooled specificity of MRI compared to template guided biopsy was 37% (CI 29%-46%)[128]. The specificity has been higher in other meta-analyses, for example 79% (CI 0.68-0.86)[149], but these had less stringent inclusion criteria and included studies with inferior reference standards. The low specificity of MRI is caused by several benign entities that produce MRI signal abnormalities yielding false-positive results. These include prostatitis, glandular benign prostatic hyperplasia, post-inflammatory glandular atrophy and fibrosis[168].

False positives are a common problem for image-based screening tests. LDCT is a highly sensitive test for lung cancer but also has a low specificity due to the high detection of non-calcified benign nodules[169]. Breast MRI has been reserved for high-risk women because it has low specificity and a high cost compared to mammography[170]. Even highly specific screening tests will generate a large number of false-positive results across multiple screening cycles.

Prostate MRI is reported using a 5-point suspicion scale and a score of 3 is defined as equivocal for the presence of clinically significant cancer. The current specificity of MRI is based on a lesion being classified as suspicious using a cut off score ≥3. However, it is not universally accepted that this cut-off is the optimum threshold to trigger a biopsy and other clinical parameters such as PSA density are increasingly utilised to risk stratify equivocal scan results for biopsy[171].

Indeed, in a screening population it is likely that a higher cut-off score, for instance of 4 or even 5, would be preferred as it might offer a more acceptable balance between sensitivity

and specificity. This has been explored in a secondary care population where a meta-analysis of 15 studies reported that a score ≥3 yielded 95% sensitivity and 47% specificity, but a score ≥4 yielded a better balance with 89% sensitivity and 74% specificity[146]. This higher threshold provides additional advantages as there is a strong correlation between cancer detection and MRI score with detection rates rising from 12%, 60% to 83% for MRI scores of 3, 4 and 5, respectively[133].

3.4.4.2 Risk of missing low-volume significant prostate cancer

Factors which improve specificity may harm the sensitivity of MRI and increase the risk of missing significant cancers. At present, the majority of prostate cancers not identified by MRI are organ-confined Gleason 3+3 (ISUP GG 1) which is regarded as a positive attribute for MRI[129]. There remains a small risk of missing significant disease particularly if it is low volume[172], located at the apical or dorsolateral segments[141] or if it is histologically diffuse intermixed with benign prostatic tissue[173]. There are also reports of rare non-acinar subtypes such as ductal or mucinous carcinoma which may be missed by MRI[127].

It seems unlikely that any screening modality will eliminate all false negative results while the definition of significant disease incorporates low volume Gleason 3+4. When the diagnostic accuracy of MRI has been evaluated against tumour volume in radical prostatectomy specimens, MRI could correctly identify over 94% of significant tumours above 0.5cm^3 but this falls to only 26% for tumours less than 0.5 cm^3 in the peripheral zone[172]. In historical definitions of insignificant disease, tumours of this volume would have been classified as insignificant. The original definition proposed by Stamey et al[174] included all tumour volumes less than 0.5cm^3 and this was extended to 1.3 cm^3 for Gleason 3+3 in a study within the ERSPC cohort[175].

It remains contentious whether any mortality benefit would be conferred in identifying such low volume Gleason 3+4 disease. Prior to 2005, tumours with small volumes of pattern 4 (<5%) were classified as Gleason 3+3 with tertiary pattern 4. Following revisions to the Gleason grading system which removed tertiary patterns on biopsy these are now graded as Gleason 3+4 (ISUP GG 2)[176]. So there is a subset of patients with favourable Gleason 3+4 disease[177] and this is supported by evidence from active surveillance programmes where men with non-visible Gleason 3+4 have a lower risk of progression on active surveillance compared to men with MRI-visible Gleason 3+4[142].

The risk of MRI missing clinically significant cancer has been improved with the release of Prostate Imaging Reporting and Data System (PI-RADS) v2[173] which sets the capabilities of MRI as able to identify Gleason score > 7 (including 3+4 with prominent but not predominant Gleason 4 component) and/or tumour volume ≥0.5ml. A validation study evaluated PI-RADSv2 against 150 prostatectomy specimens and concluded that while MRI identified ≥94% of significant foci ≥0.5 mL, it had limited capacity at lower volumes and only identified 26% of significant peripheral zone tumours at volumes ≤0.5mL[172]. It seems unlikely that any screening modality will entirely eliminate false negative results.

3.4.4.3 Incidental findings

Incidental findings are abnormalities identified by the screening test which are unrelated to the condition being screened. It is a common problem across image-based screening tests and it can pose a challenge as the majority of findings are likely insignificant but may trigger additional costly and burdensome confirmatory investigations[178]. The problems of incidental findings have been seen in screening CT colonoscopy where it is estimated over a quarter of screened subjects had extracolonic findings[178], and in LDCT for lung cancer where 15% of scans have incidental non-pulmonary findings requiring further testing[179].

There is limited evidence regarding the prevalence of incidental findings during MRI for prostate cancer. It is known that incidental findings are most common in imaging tests which have a large field of view (FOV). The standard FOV in prostate MRI is narrow and extends 12-20cm including the prostate, seminal vesicles, bladder, rectum and femoral heads[180]. There is the possibility of identifying a range of incidental abnormalities such as rectal cancers, bladder cancers or femoroacetabular abnormalities.

3.4.4.4 Reproducibility

Another important criterion for a screening test is that it must provide consistent results when performed across diverse centres and interpreted by different clinicians. The rapid advances and widespread uptake of MRI has meant that there is a lack of standardisation between centres and variability in the quality of acquisition and reporting[129]. There are continuing efforts to further improve the standard image acquisition techniques and reporting protocols with the release of PI-RADSv2.1[181] and a UK consensus for Likert scoring[182]. The interobserver variability has been moderate with a kappa coefficient of 0.45 for PI-RADS ≥3[183], similar to the level of agreement achieved during the early phase of screening mammograms (kappa 0.47)[184]. If the cut-off score for a positive MRI is increased

to ≥4, the kappa co-efficient of MRI is equivalent to current mammography practice at 0.55[183].

3.4.4.5 Safety

Even a minor side effect when accumulated across the population will significantly offset any benefit from a screening test. The main risk of a standard mpMRI is the gadolinium-based contrast agent which carries a small risk of contrast induced nephropathy and medical cover is often required on site in case of anaphylaxis which adds a further financial burden to healthcare systems. There is also concern regarding gadolinium accumulation in the brain[185].

3.4.4.6 Acceptability

There are potential issues with the acceptability of MRI as a screening test and it will be important to establish the influence of MRI on uptake rates within a screening programme. PSA is generally seen as an acceptable screening test and the Cluster Randomised Trial of PSA Testing for Prostate Cancer (CAP) has reported an uptake of about 40% for a organised PSA screening programme[186]. This is lower than breast or cervical screening which have uptake rates consistently above 70%[187] but these are in the context of an national screening programme with extensive resources, advertising and public acceptance. CaP also used a randomised consent (Zelen) design which is known to have lower acceptance rate once the participants in the intervention arm are informed of their allocation. The ERSPC study also showed this with countries using the Zelen design having 62-68% acceptance intervention whereas those using patient-level consent to randomisation showing compliance to intervention of 88-100%. There is a risk of reduced uptake due to lack of availability and accessibility of MRI requiring longer travelling times compared to a serum biomarker which can be taken by any health professional. Certain groups may have difficulty tolerating the scan particularly patients with anxiety, claustrophobia or those unable to be immobile for the duration of the scan. There is the option of offering individuals with anxiety or claustrophobia an anti-anxiolytic before the MRI.

3.4.4.7 Cost Effectiveness

Introducing a population-based screening programme presents a huge resource commitment and requires a large infrastructure to deliver an effective programme. The most sensitive test is not necessarily selected if it is not cost-effective. A key challenge for MRI as a screening test is the high operational cost and large capital investment. From a UK

healthcare perspective, the cost from the NHS tariff would be £108 per scan including reporting[188] in comparison to a PSA test at £4[189]. This can be compared to mammography at £33.50[190] and LDCT for lung cancer at £69[188]. These prices are obviously determined by the healthcare setting where screening programmes are delivered at scale and MRI costs are currently only applicable to secondary care.

The cost effectiveness of a screening programme is determined by multiple factors beyond the unit cost of a screening test. Factors such as the screening interval, reductions in subsequent investigations or overdiagnosis can balance out a high cost test. The cost of MRI already compares favourably to flexible sigmoidoscopy for colorectal cancer at £304[188] which has been introduced across certain European countries[191-193]. Colonoscopy screening offers an even higher level of sensitivity at a higher cost of £429[188] and has been instigated in the USA[194] and Poland[195]. These approaches have been proven cost-effective as their high diagnostic accuracy means they can be repeated at less frequent intervals compared to cheaper stool based tests[196].

3.4.4.8 Capacity

Screening tests need to be widely available and accessible to the target population to ensure adequate uptake. It is likely that the main barrier to MRI screening is currently the limited availability of MRI machines and trained personnel to deliver the service. The feasibility of delivering even routine pre-biopsy prostate MRI has been questioned in the secondary care population[139]. A lack of endoscopy capacity has been a limiting factor in the roll out of the flexible sigmoidoscopy screening programme for colorectal cancer. However, a current lack of capacity should not be a reason to prevent the evaluation of MRI screening in clinical trials. If MRI screening is shown to be cost-effective then further capital investment in dedicated screening units comparable to breast screening will be required. The implementation of an MRI screening programme would require a pilot period prior to a phased roll out, to ensure that any capacity and workforce issues are properly managed.

3.5 Discussion

3.5.1 Principle findings

Screening tests can have a large impact on public health and careful evaluation is required prior to their introduction. This chapter has presented the key literature relevant to MRI as a screening test and yielded insight in two specific areas. First, the potential benefits associated with MRI as a screening test and second, the challenges and limitations likely to be associated with MRI screening.

For the potential benefits of MRI, five themes emerged related to the high diagnostic accuracy and low invasiveness of MRI. The current literature shows MRI can reliably identify clinically significant disease ≥0.5cm^3 in volume with a high sensitivity and negative predictive value[88]. If the performance characteristics are maintained in a screening setting, MRI has the potential to improve detection rates and reduce overdiagnosis which have been the primary issues with PSA-based screening programmes.

For the potential problems associated with standard MRI, I describe eight themes which included risk of lower specificity, missing low-volume disease, incidental findings, reproducibility, safety, acceptability, higher costs and capacity. In the subsequent chapters of my thesis it will be important to weight any benefits associated with MRI against these potential harms.

3.5.2 Implications of findings

The number of potential problems with standard MRI highlight the need for the test to be adapted to be made suitable for a screening population. A screening MRI protocol needs to be simple and practical without significantly impacting the accuracy of the test. The standard diagnostic MRI incorporating T2w, DWI and DCE sequences is expensive and requires up to 40 minutes including patient preparation and administration of contrast[197].

There have already been efforts to streamline MRI by removing MR spectroscopy and endorectal coils. Further progress is needed to improve image acquisition time and cost-effectiveness for screening. Recent studies have described fast bi-parametric MRI (bp-MRI) protocols incorporating non-contrast T2W and DWI sequences to achieve an image acquisition time of 15 minutes at significantly reduced cost[198].

If these tests are to be suitable for screening they will need to be evaluated beyond a diagnostic population to the general population of men in which they are intended to be used. This is because the performance of a test can be overestimated by evaluating it in a sample which is not representative of the population it will be applied to. This effect was first described by Ransohh and Feinstein[199] who observed that test performance varied among different population sub-groups and rarely performed to the same level in the general population.

In prostate cancer, an extreme form of this bias occurs in studies which validate diagnostic tests for prostate cancer against radical prostatectomy specimens. These studies include only the sub-group of men who had a diagnosis of cancer on biopsy and subsequently chose to have surgery often on the basis of sufficiently high grade and volume disease to warrant radical prostatectomy. Men with less advanced disease which may be more difficult to diagnose, although may still warrant treatment, are excluded in such study designs. By omitting low volume significant disease, this can lead to low rate of false negatives and underestimation of sensitivity and NPV of MRI.

Given that the performance of a diagnostic test is so dependent on the risk profile of the population, researchers must report sufficient information to allow the effect of spectrum bias to be assessed. The STARD statement for reporting studies of diagnostic accuracy[200] requires authors to describe the study population in detail. This was not adhered to in all the studies evaluated for this narrative review. The chapters for my thesis will adhere to this guidance, where relevant, by reporting how, where and when participants were identified, the eligibility criteria for selection and whether participants formed a consecutive, random or convenience series.

3.5.3 Limitations

The main limitation of this chapter is the narrative design which is more likely to introduce bias compared to a systematic review. A literature search for the purposes of a systematic review was conducted but identified only a single pilot study investigating MRI as an independent screening test. This lack of empirical evidence precluded the original plan for a systematic review +/- meta-analysis and highlighted the need for further research in this area. Instead, a narrative review was conducted to synthesise a broader range of themes across the literature although this design does not allow definitive conclusions to be drawn.

3.6 Conclusion

The lack of studies which directly investigate MRI as a first-line screening test makes it difficult to draw any firm conclusions about its potential for screening. The existing literature has focused on MRI in diagnosing prostate cancer in secondary care where it plays a pivotal role now as a result of its reported high sensitivity for the detection of clinically significant cancer. While these performance characteristics could be attractive for a screening test, there are many inherent challenges and complexities which need to be considered when evaluating MRI as a screening test. This review has helped identify the current gaps in the literature which will need to be addressed in subsequent chapters. Importantly, implementation of an MRI screening programme requires a shorter MRI protocol and this will be evaluated in the subsequent chapter.

Chapter 4 – Validation of a fast MRI protocol for the detection of clinically significant prostate cancer

4.1 Overview

The standard diagnostic multi-parametric MRI requires up to 40 minutes' examination time including patient preparation and administration of intravenous contrast[197]. In this chapter, I seek to mitigate some of the challenges highlighted in Chapter 3 by evaluating the performance of an MRI protocol with a shorter examination time. This Chapter forms the basis of work published in BJU International[iii].

4.2 Introduction

The standard MRI recommended by the recent PI-RADS version 2.1 guidelines incorporates T2-weighting, diffusion weighting and dynamic contrast enhanced sequences (DCE) and is referred to as multiparametric MRI (mpMRI)[201]. This protocol was not developed for screening and instead was designed for a diagnostic setting focusing on detection, localisation and local staging of significant prostate cancer. The protocol was developed with the primary goal of maximising diagnostic accuracy with less prominence placed on costs, capacity, time, logistics and side effects of intravenous contrast.

Diagnostic accuracy is an important feature of a screening test but there are additional considerations such as safety, cost effectiveness and minimising overdiagnosis. The role of DCE requiring administration of a gadolinium-based contrast agent has been particularly controversial[202]. There are rare but serious side effects including anaphylaxis/anaphylactoid reactions, nephrogenic systemic fibrosis and gadolinium intracranial deposition which are important considerations for an MRI screening test that could be applied over a large population who are predominately healthy.

The role of DCE has been receiving less importance with time. In earlier versions of PI-RADS, there was a similar level of importance given to each T2W, DWI and DCE sequence without emphasising any dominant sequence[203]. The recent version of PI-RADSv2 has limited the role of DCE to a binary score which can only differentiate indeterminate lesions in the peripheral

[iii] Eldred-Evans, D., et al. (2020). "Added value of diffusion-weighted images and dynamic contrast enhancement in multiparametric magnetic resonance imaging for the detection of clinically significant prostate cancer in the PICTURE trial." BJU International **125**(3): 391-398.

zone (PZ)[201]. It also recommends dominant sequences based on zonal anatomy so that DWI is dominant in the PZ and T2W is dominant in the transition zone (TZ).

Further, DCE increases the time to acquire and interpret an MRI. Most DCE protocols require an additional 15 minutes per patient accounting for intravenous access compared to 8-10 minutes for DWI and 11 minutes for T2 alone with axial and coronal views. There have been reports that with increased radiologist experience and improvements in MRI technology it may be feasible to reduce the number of mpMRI sequences[123, 204, 205].

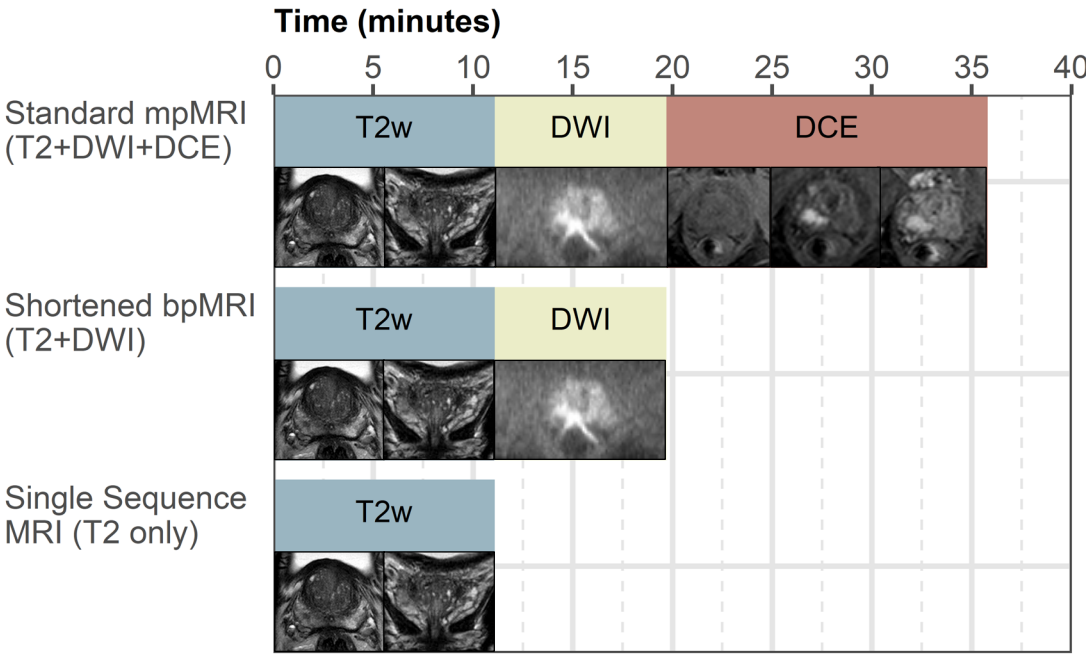


Figure 5: Time saving from proposed shorter MRI protocols as compared to the standard mpMRI. The addition of DCE and the need for IV access increases time ~15 minutes. The addition of DWI increases study time ~8 mins.

If it can be shown that specific sequences can be safely omitted without impacting diagnostic accuracy, this could address the limited availability of MRI which will challenge the use of MRI for screening. The high demand for screening will mean that the examination must be made shorter and more cost efficient. However direct comparison of these different sequences for predicting clinically significant prostate cancer has yielded conflicting results. Previous systematic reviews have shown that a shortened bpMRI was less sensitive than standard mpMRI[137]. However, recent paired cohort studies have suggested that DCE may not be necessary[123, 204, 205].

Therefore the principal objects of this study were (1) to compare diagnostic accuracy of each sequence (2) to identify the sequences suitable for screening considering the balance of

biopsy rates, diagnostic accuracy and overdiagnosis rates; and (3) to compare interobserver variability between each sequence.

4.3 Methods

4.3.1 Study Design

This was a paired-cohort validating confirmatory study designed to provide level 1 evidence on the diagnostic accuracy of each MRI sequence. The study uses data from the PICTURE trial[iv] which was approved by the National Research Ethics Committee London (reference 11/LO/1657) and the full trial protocol has been previously published[206]. This study was written according the Standards for Reporting Diagnostic Accuracy guidelines for reporting diagnostic accuracy studies[207].

4.3.2 Participants

Participants were recruited in urology outpatient clinics and were eligible for enrolment if they had undergone prior TRUS-biopsy and been advised as part of standard of care to undergo a repeat biopsy for further risk stratification. Participants were excluded if they had previous treatment for prostate cancer; a contraindication to MRI or artefact which would reduce MRI quality; a prostate gland volume >/=80ml or other medical conditions meaning they were unable to have general or regional anaesthesia.

4.3.3 Procedures

4.3.3.1 MRI protocol

mpMRI was performed using a 3T scanner with a pelvic-phased array coil. All MRIs were compliant with European Society of Uroradiology guidelines[208]. The acquisition protocol consisted of T1W and T2W sequences, DWI with high b-value (b=2000) and apparent diffusion coefficient (ADC) map using multiple b-values (b=0,150,500,1000) and DCE with gadopentetate dimeglumine (Magnevist, Bayer AG, Leverkusen, Germany). The full acquisition parameters and time for each sequence are listed in Table 6.

[iv] Disclosure: The PICTURE trial had been completed prior to commencement of this thesis. This chapter was a sub-analysis of data which had been previously collected. I performed the statistical analysis for this chapter and wrote the manuscript for publication in the BJU International

Table 6: MRI Sequences in PICTURE

Sequence	Coil	TR	TE	FA	WFS (pix)	BW Hz/Px	FoV mm	Thick-ness	Gap	FSE/TSE factor	Phasing direction	FS	ACQ	TFE Shots	Interval (ms)	Scan Duration
T2 TSE coronal	Dual	6128	100	90	2.704	160.7	180	3	3	16	R > L	No	300 × 290			05:55.4
T2 TSE axial	Dual	5407	100	90	2.704	160.7	180	3	0	16	R > L	No	300 × 290			05:13.6
T2 sag	Dual	1579	100	90	1.999	217.3	240	5	5	20	A > P	No	120 × 89			00:18.9
T1W TSE	Sense Torso	487	8.0	90	1.997	217.6	240	3	3	4	R > L	No	184 × 184			03:06.8
VISTA sense	Dual	2000	200	90	1.108	392.0	200	3	3	66	R > L	No	248 × 187			04:26.0
DWI 0 150 500 1000	Dual	2753	80	90	40.353	10.8	220	5	0		A > P	SPAIR	168 × 169			05:16.5
DWI sFOV	Dual	2824	89	90	23.048	18.9	90	5	0		A > P	SPIR	68 × 61			05:33.2
DWI b2000	Dual	2000	78	90	44.108	9.9	220	5	0		A > P	SPIR	168 × 169			03:40.0
DCE 2 dyn mod sense	Dual	5.8	2.8	10	1.766	246.1	180	3	0	38	R > L	SPAIR	140 × 177	49	280	00:28.9

Abbreviations: Repetition Time (TR), Echo Time (TE), Flip Angle (FA), Water-Fat Shift (WFS), Bandwidth (BW), Field of View (FoV),

4.3.3.2 MRI Reporting

MRI scans were reported in a sequential blinded fashion so that three sets of reports were generated; T2W alone, T2W+DWI and T2W+DWI+DCE. Reporting was completed in a single session with each sequential report locked after being issued. The reporting was completed by a board-certified radiologist with over 10 years of experience in prostate MRI interpretation and reporting a high volume of MRIs per annum (>1500 scans/year). To assess interobserver agreement, 50 (20%) of mpMRI consisting of all sequences were randomly selected for re-reporting by a second expert radiologist blinded to the original reports.

A 5-point Likert Assessment System was used to rate the likelihood of clinically significant disease as highly unlikely (1), unlikely (2), equivocal (3), likely (4) or highly likely (5). This scoring system has been prospectively validated in the PROMIS trial[88] and has been recommended for use by the UK national consensus meeting[209]. Comparative studies have shown that it provides similar results to PI-RADS [148, 210].

4.3.3.3 Reference Test

All participants underwent template mapping biopsy irrespective of the findings on mpMRI. Participants were blinded to the mpMRI results to reduce selection bias and increase compliance with TPM biopsy. The biopsy mapping procedure followed a pre-defined protocol in which biopsy cores were taken every 5-10mm using a brachytherapy grid. Additional targeted cores could also be taken within the biopsy protocol which were processed and reported separately. As per the original analysis plan and the primary outcomes of mpMRI validation [211] only the TPM biopsies were used in this report. All biopsies were reported in accordance with the 2005 International Society of Urological Pathology (ISUP) recommendations [40] by one of two expert uropathologists blinded to the mpMRI report. Any negative biopsy was double reported as part of a quality control process.

4.3.4 Outcomes

The definition for a suspicious MRI to determine biopsy was evaluated on two thresholds; Likert ≥ 3 and Likert ≥ 4. The definition of clinically significant disease was set using validated criteria which have been developed for TPM biopsy[41]. The primary definition (UCL/Ahmed definition 1) was the presence of dominant Gleason pattern 4 or greater and/or a cancer core length (CCL) involvement of >/=6mm of any Gleason score. The secondary definitions were a) UCL/Ahmed 2 (any Gleason score 7 and/or CCL involvement of >/=4mm of any Gleason score), b) any Gleason >/=3+4 and c) any Gleason >/=4+3.

4.3.5 Statistical Analysis

The analysis conducted in this chapter was a secondary objective using data from the PICTURE trial. The primary objective of the PICTURE trial was pre-defined and set to evaluate the negative predictive value (NPV) of mpMRI using a precision-based estimate. The sample size calculation was performed for this pre-defined primary endpoint prior to the start of the trial[206].

Descriptive statistics were used for baseline characteristics, distribution of Likert scores and cancer detection rates. Sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) were calculated for each of the three sequences with binomial 95% confidence intervals, with a score of 3 or above or a score 4 or above designating a suspicious/positive MRI. Receiver–operator curves (ROC) were constructed and DeLong's test was used to compare the area under the ROC (AUROC)[212].

The interobserver agreement was calculated using weighted kappa and proportion of agreement and assessed using AUROC. The weighted kappa allows for the magnitude of the disagreements to be taken into account. The weighting system used resulted in the weights 0.75, 0.5, 0.25 and 0 for MRI ratings scores that differed by 1, 2, 3 and 4 respectively. All analyses were conducted using R Version 4.0.2[213] (R Foundation for Statistical Computing, Austria) with any tests of significance using two-sided p=0.05 as the threshold for statistical significance.

4.4 Results

4.4.1 Study Population

A total of 330 participants were enrolled into the study and there were 81 withdrawals. A study flow chart is provided in Figure 6 which illustrates the outcomes of the 249 participants included in this analysis. During the study period, all participants underwent TPM biopsy where a median (IQR) 49 (40-50) cores were taken. Any cancer was detected on TPM-biopsy in 209 (83.9%) of 249. Using definition 1, clinically significant cancer was detected on TPM-biopsy in 103 (41%) of 249.

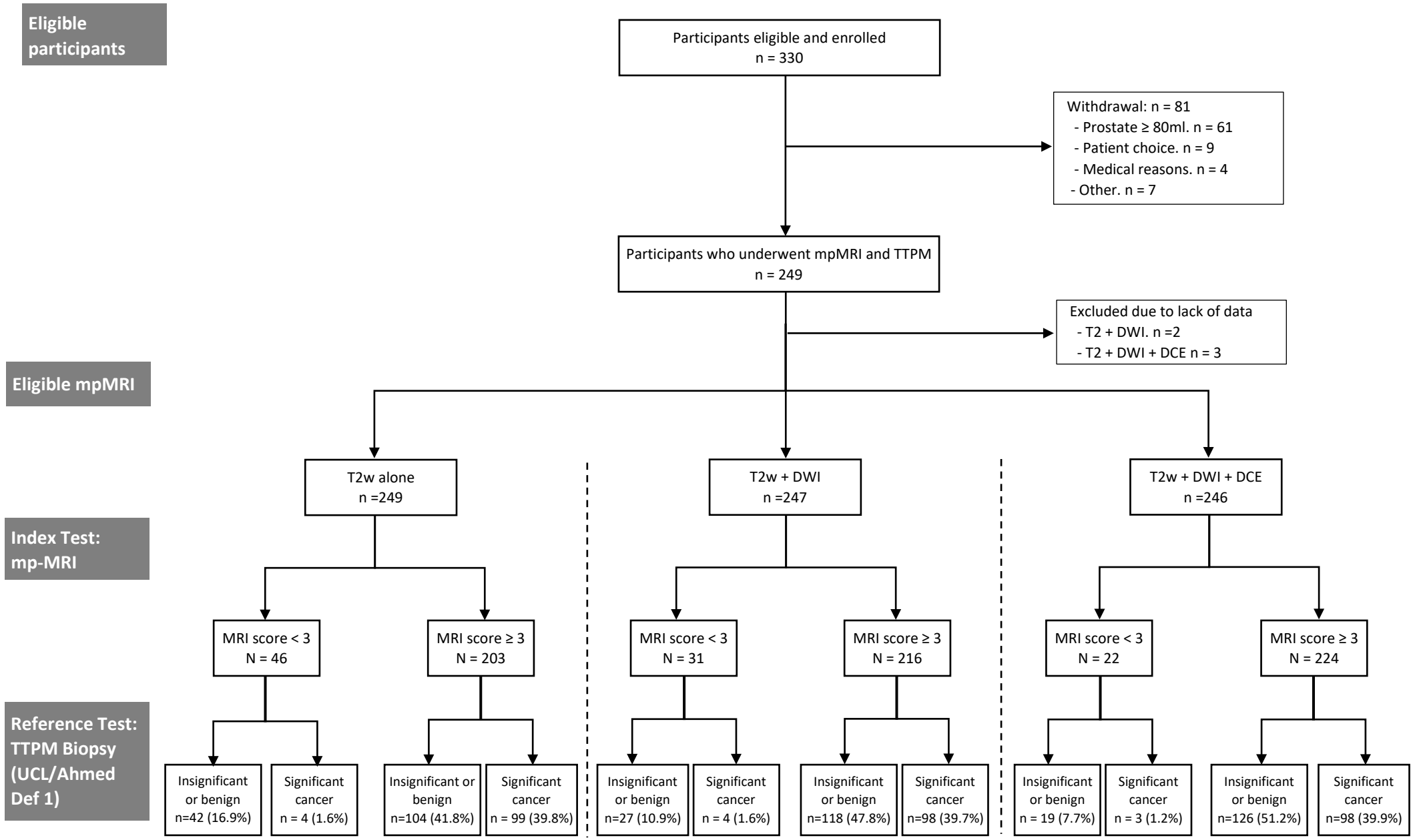
Participants had a mean age 62.0 (SD 7.0) years and median PSA 6.8ng/ml (IQR 4.8-9.8). All participants had a previous TRUS biopsy which had shown no cancer in 76 (30.5%), Gleason 6 in 121 (48.6%) and low volume Gleason 7 in 52 (21.1%). Further baseline characteristics for 249 participants are shown in Table 7.

Table 7: Characteristics of Participants

	Participants (n = 249)
Age, years (SD)	62 (7.0)
PSA (ng/ml), Median (IQR)	6.8 (4.8-9.8)
Prostate volume, median (IQR)	37.0 (26.8-50.0)
Previous biopsies, median (IQR)	1 (1-2)
Previous TRUS biopsy result (%)	
Benign	76 (30.5)
Gleason 3+3	121 (48.6)
Gleason ≥ 3+4	52 (21.1)
TPM biopsy outcomes (%)	
Benign	111 (27.2)
Gleason score	
3+3	66 (26.5)
3+4	110 (44.2)
4+3	29 (11.7)
4+4	3 (1.2)
3+5	1 (0.4)
MCCL, median (IQR)	4 (2-7)
Cores positive for cancer, median	6 (2-11)

Abbreviations: MCCL = Maximum cancer core length

Figure 6: Standards for reporting of diagnostic accuracy (STARD) flowchart



4.4.2 MRI scores and biopsy rates

The distribution of Likert scores is visualised in Figure 7. The T2 only sequence had more equivocal Likert 3 lesions (44.7% (95% CI 38.4-51.2)) compared to the other sequences (T2+DWI 32.5% (95% CI 26.8-38.8) / T2+DWI+DCE 33.3% (95% CI 27.5-39.6, p<0.001). There were more Likert 5 lesions for T2+DWI (30.9% (95% CI 25.3-37.1)) and T2+DWI+DCE (37.8% (95% CI 31.8-44.2)) compared to T2-only (11.8% (95% CI 8.2-16.6), p<0.01).

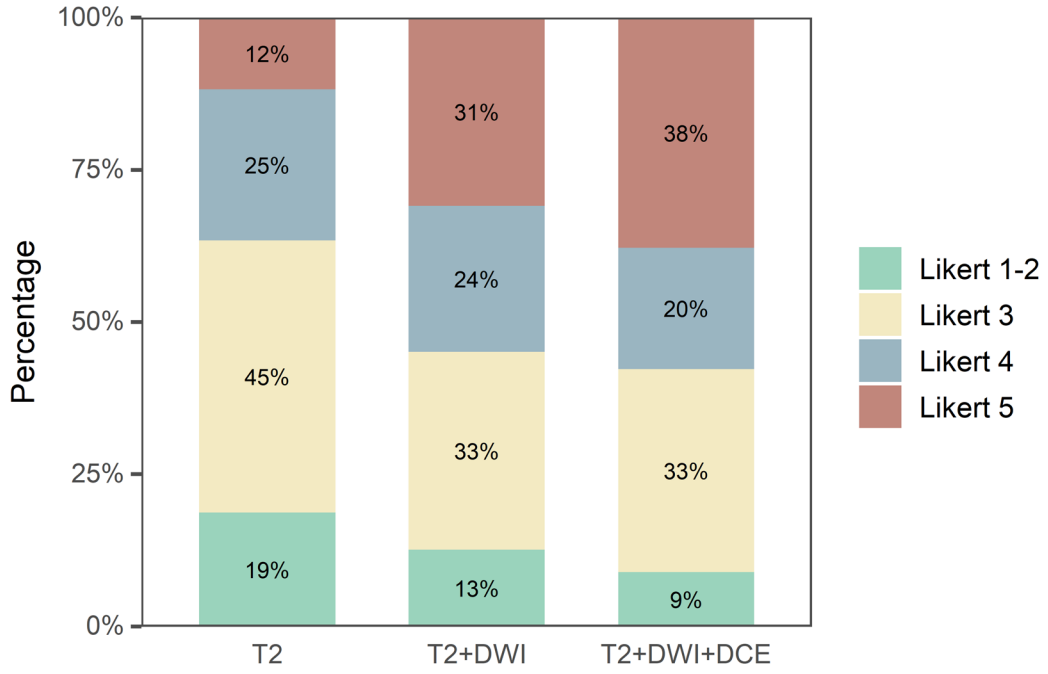


Figure 7: Distribution of Likert scores for each MRI sequence.

A comparison between biopsy rates showed that if a threshold of Likert ≥ 3 was utilised, the biopsy rate for T2 only was 81.3% (95% CI 75.7-85.9), T2+DWI was 87.4% (95% CI 82.4-91.2) and T2+DWI+DCE was 91.1% (95% CI 86.6-94.2). There was no significant difference in biopsy rate between either T2 and T2+DWI (p = 0.082) or T2+DWI and T2+DWI+DCE (p=0.245). There was a net increase in biopsy rate of 9.8% between T2 and T2+DWI+DCE (p = 0.003).

At Likert ≥ 4, the biopsy rate for T2 only was 36.6% (95% CI 30.6-43.0), T2+DWI was 54.9% (95% CI 48.4-61.2) and T2+DWI+DCE was 57.7% (95% CI 51.3-63.9). There was a net increase in biopsy rate between T2 and T2+DWI (+18.3% 95% CI 9.5-26.7, p<0.001) and between T2 and T2+DWI+DCE (+21.1%, 95% CI 12.3-29.4, p<0.001). There was no difference in biopsy rates between T2+DWI and T2+DWI+DCE (+2.8%, 95% CI -5.9-11.5, p=0.586).

4.4.3 Diagnostic accuracy (Likert ≥ 3)

For definition UCL/Ahmed 1, T2W alone had a sensitivity of 96% (95% CI 90–99), a specificity of 29% (95% CI 22–37), a PPV of 49% (95% CI 42–56) and an NPV of 91% (95% CI 79–98). For T2W+DWI, sensitivity was 96% (95% CI 90–99), specificity 29% (95% CI 22–37), PPV 49% (95% CI 42–56) and NPV 91% (95% CI 79–98). For T2W+DWI+DCE, sensitivity was 96% (95% CI 90–99), specificity 29% (95% CI 22–37), PPV 49% (95% CI 42–56) and NPV 91% (95% CI 79–98). Overall accuracy for definition 1, as assessed by area under the receiver operating characteristic curve (AUROC), for T2W, T2W+DWI and T2W+DWI+DCE were 0.74 (95% CI 0.68-0.80), 0.76 (95% CI 0.71-0.82) and 0.77 (95% CI 0.71-0.82), respectively (p=0.55)

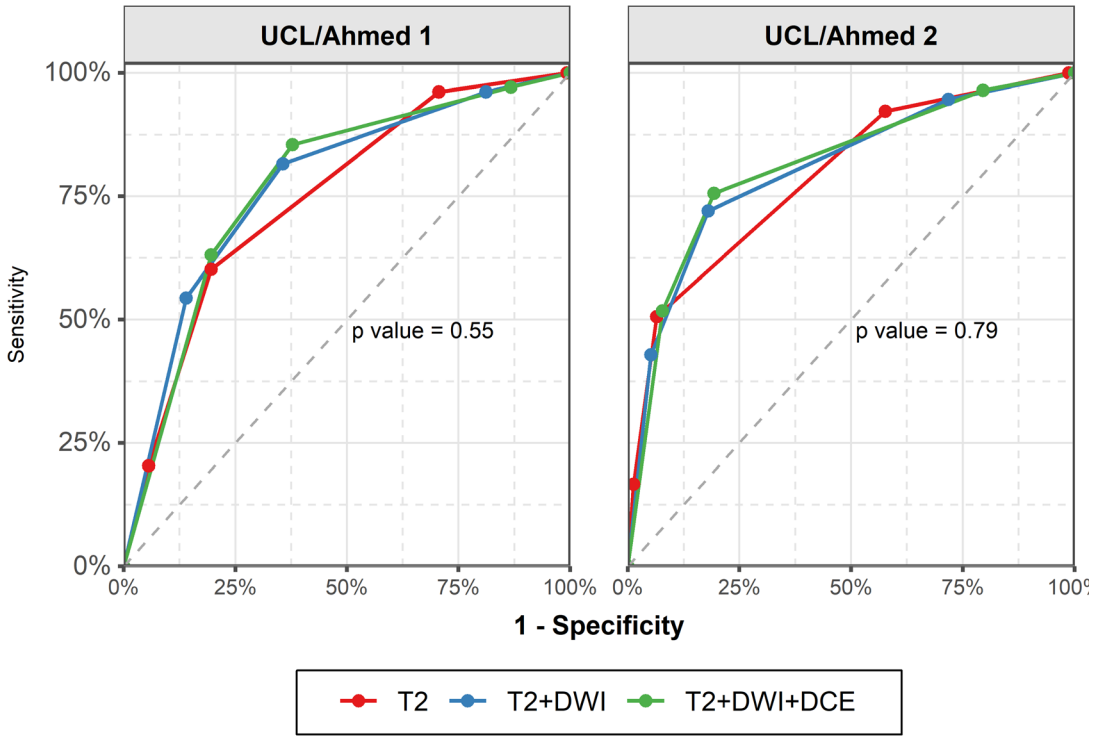


Figure 8: Comparison of area under the receiver-operating characteristic curve for UCL/Ahmed definitions 1 and 2 for clinically significant cancer.

For definition UCL/Ahmed 2, T2W alone had a sensitivity of 92% (95% CI 87-96), a specificity of 41% (95% CI 30-52), a PPV of 76% (95% CI 70-82), and an NPV of 72% (95% CI 57-84). For T2W+DWI, sensitivity was 95% (95% CI 90-98), specificity 27% (95% CI 18-38), PPV 73% (95% CI 66-79) and NPV 71% (95% CI 52-86). For T2W+DWI+DCE, sensitivity was 96% (95% CI 92-99), specificity 20% (95% CI 12-30), PPV 71% (95% CI 65-77) and NPV 73% (95% CI 50-89). Overall accuracy for definition 2, as assessed by area under the AUROC, for T2W, T2W+DWI and T2W+DWI+DCE were 0.77 (95% CI 0.71-0.83) 0.78 (95% CI 0.72-0.84) and 0.79 (95% CI 0.73-0.84), respectively (p=0.79).

For the definition Gleason ≥3+4, T2W alone had a sensitivity of 93% (95% CI 88-97), a specificity of 34% (95% CI 25-44), a PPV of 66% (95% CI 59-72) and an NPV of 78% (95% CI 64-89). For T2W+DWI, sensitivity was 95% (95% CI 90-98), specificity 23% (95% CI 15-32), PPV 63% (95% CI 56-69) and NPV 77% (95% CI 59-90). For T2W+DWI+DCE, sensitivity was 97% (95% CI 92-99), specificity 16% (95% CI 10-24), PPV 61% (95% CI 54-67) and NPV 77% (95% CI 55-92). Overall accuracy for T2W, T2W+DWI and T2W+DWI+DCE were 0.72 (95% CI 0.66-0.78), 0.73 (95% CI 0.67-0.79) and 0.74 (95% CI 0.68-0.80) , respectively (p=0.53).

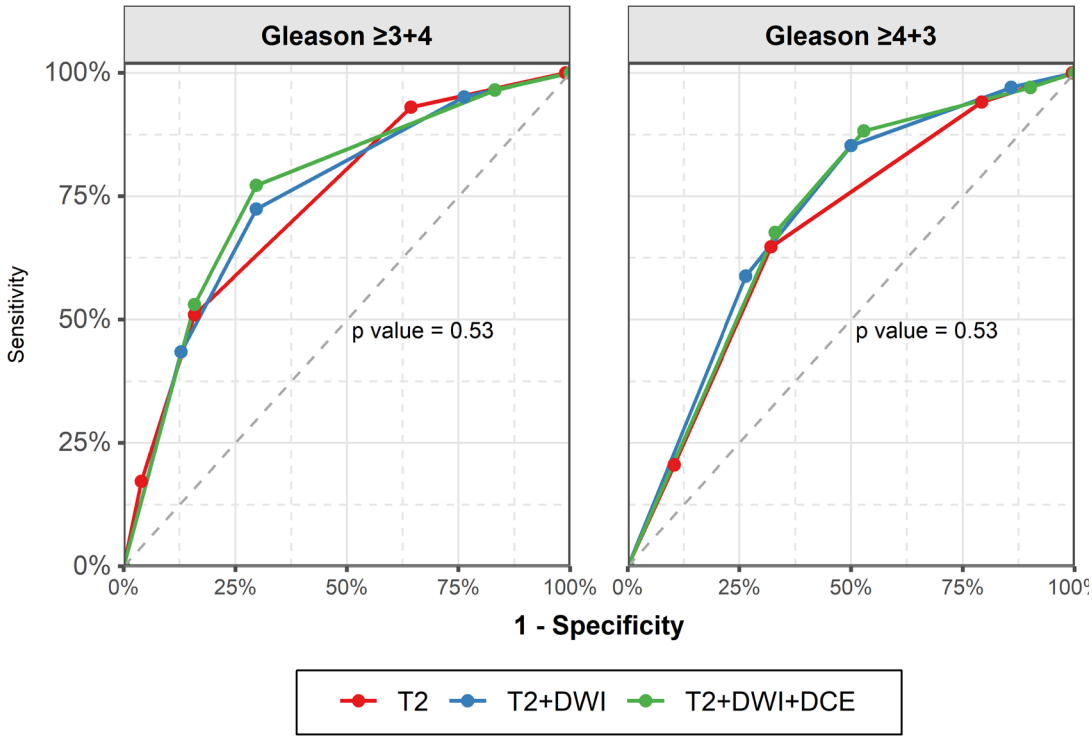


Figure 9: Comparison of area under receiver operating characteristic curve for Gleason ≥ 3+4 and Gleason ≥ 4+3.

For the definition Gleason ≥4+3, T2W alone had a sensitivity of 94% (95% CI 79-99), a specificity of 20% (95% CI 15-26), a PPV of 15% (95% CI 10-20) and an NPV of 85% (95% CI 96-100). For T2W+DWI, sensitivity was 97% (95% CI 84-100), specificity 14% (95% CI 10-19), PPV 14% (95% CI 10-20) and NPV 97% (95% CI 83-100). For T2W+DWI+DCE, sensitivity was 97% (95% CI 84-100), specificity 10% (95% CI 6-15), PPV 14% (95% CI 10-19) and NPV 77% (95% CI 96-100). Overall accuracy for T2W, T2W+DWI and T2W+DWI+DCE were 0.68 (95% CI 0.59-0.77), 0.71 (95% CI 0.62-0.80) and 0.71 (95% CI 0.63-0.79), respectively (p=0.53)

Table 8: Diagnostic accuracy of each MRI sequences for each definition of clinically significant prostate cancer at threshold ≥ 3

	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
UCL/Ahmed 1: Dominant Gleason pattern 4 or greater and/or a cancer core length (CCL) involvement of ≥6mm of any Gleason score (Prevalence 42%)				
T2W	96 (90-99)	29 (22-37)	49 (42-56)	91 (79-98)
T2W + DWI	96 (90-99)	19 (13-26)	45 (39-52)	87 (70-96)
T2W + DWI + DCE	97 (92-99)	13 (8-20)	44 (37-51)	86 (65-97)
UCL/Ahmed 2: Any Gleason pattern 7 or greater and/or a cancer core length involvement of ≥4mm of any Gleason score (Prevalence 68%)				
T2W	92 (87-96)	41 (30-52)	76 (70-82)	72 (57-84)
T2W + DWI	95 (90-98)	27 (18-38)	73 (66-79)	71 (52-86)
T2W + DWI + DCE	96 (92-99)	20 (12-30)	71 (65-77)	73 (50-89)
Any Gleason score ≥3+4 (Prevalence 59%)				
T2W	93 (88-97)	34 (25-44)	66 (59-72)	78 (64-89)
T2W + DWI	95 (90-98)	23 (15-32)	63 (56-69)	77 (59-90)
T2W + DWI + DCE	97 (92-99)	16 (10-24)	61 (54-67)	77 (55-92)
Any Gleason score ≥4+3 (prevalence 14%)				
T2W	94 (79-99)	20 (15-26)	15 (10-20)	85 (96-100)
T2W + DWI	97 (84-100)	14 (10-19)	14 (10-20)	97 (83-100)
T2W + DWI + DCE	97 (84-100)	10 (6-15)	14 (10-19)	77 (96-100)

Abbreviations: PPV = Positive Predictive Value; NPV = Negative Predictive Value

4.4.4 Diagnostic accuracy (Likert ≥ 4)

At the higher threshold of Likert ≥ 4, for definition UCL/Ahmed 1, T2W alone had a sensitivity of 60% (95% CI 50-70), a specificity of 80% (95% CI 73-87), a PPV of 69% (95% CI 58-78), and an NPV of 74% (95% CI 66-80). For T2W+DWI+DCE, sensitivity was 82% (95% CI 73-89), specificity 64% (95% CI 56-72), PPV 49% 62% (95% CI 73-70), and NPV 83% (95% CI 75-89). For T2W+DWI+DCE, sensitivity was 85% (95% CI 77-92), specificity 62% (95% CI 54-70), PPV 62% (95% CI 53-70), and NPV 86% (95% CI 77-92).

At a definition UCL/Ahmed 2, T2W alone had a sensitivity of 51% (95% CI 43-58), a specificity of 94% (95% CI 86-98), a PPV of 94% (95% CI 88-98), and an NPV of 47 (95% CI 39-55). For T2W+DWI, sensitivity was 72% (95% CI 65-79), specificity 82% (95% CI 72-90), PPV 90% (95% CI 83-94), and NPV 58% (95% CI 48-67). For T2W+DWI+DCE, sensitivity was 76% (95% CI 68-82), specificity 81% (95% CI 70-89), PPV 89% (95% CI 93-94), and NPV 61% (95% CI 51-70).

Table 9: Diagnostic accuracy of each MRI sequences for each definition of clinically significant prostate cancer at threshold ≥ 4

	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
UCL/Ahmed 1: Dominant Gleason pattern 4 or greater and/or a cancer core length (CCL) involvement of ≥6mm of any Gleason score (prevalence 42%)				
T2W	60 (50-70)	80 (73-87)	69 (58-78)	74 (66-80)
T2W + DWI	82 (73-89)	64 (56-72)	62 (73-70)	83 (75-89)
T2W + DWI + DCE	85 (77-92)	62 (54-70)	62 (53-70)	86 (77-92)
UCL/Ahmed 2: Any Gleason pattern 7 or greater and/or a cancer core length involvement of ≥4mm of any Gleason score (Prevalence 68%)				
T2W	51 (43-58)	94 (86-98)	94 (88-98)	47 (39-55)
T2W + DWI	72 (65-79)	82 (72-90)	90 (83-94)	58 (48-67)
T2W + DWI + DCE	76 (68-82)	81 (70-89)	89 (93-94)	61 (51-70)
Any Gleason score ≥3+4 (Prevalence 59%)				
T2W	51 (43-59)	84 (76-91)	82 (83-89)	82 (73-89)
T2W + DWI	72 (64-80)	70 (60-79)	78 (70-84)	64 (54-73)
T2W + DWI + DCE	77 (70-84)	70 (69-79)	79 (71-85)	68 (58-77)
Any Gleason score ≥4+3 (Prevalence 14%)				
T2W	65 (46-80)	68 (61-74)	24 (16-35)	92 (87-86)
T2W + DWI	85 (69-95)	50 (43-57)	21 (15-29)	95 (90-99)
T2W + DWI + DCE	88 (73-97)	47 (40-54)	21 (15-29)	96 (90-99)

Abbreviations: PPV = Positive Predictive Value; NPV = Negative Predictive Value

With Gleason ≥ 3+4 and T2 alone, there was a sensitivity of 51% (95% CI 43-59), a specificity of 84% (95% CI 76-91), a PPV of 82% (95% CI 83-89), and an NPV 82% (95% CI 73-89). For T2W+DWI, sensitivity was 72% (95% CI 64-80), specificity 70% (95% CI 60-79), PPV 78% (95% CI 70-84) and NPV 64% (95% CI 54-73). For T2W+DWI+DCE, sensitivity was 77% (95% CI 70-84), specificity 70% (95% CI 69-79), PPV 79% (95% CI 71-85), and NPV 68% (95% CI 58-77).

Finally, with a definition of Gleason ≥4+3, for T2W alone there was a sensitivity of 65% (95% CI 46-80), a specificity of 68% (95% CI 61-74), a PPV of 24% (95% CI 16-35), and an NPV 92% (95% CI 87-86). For T2W+DWI, sensitivity was 85% (95% CI 69-95), specificity 50% (95% CI 43-57), PPV 21% (95% CI 15-29), and 95% (95% CI 90-99). For T2W+DWI+DCE, sensitivity was 88% (95% CI 73-97), specificity 47% (95% CI 40-54), PPV 21% (95% CI 15-29), and NPV 96% (95% CI 90-99).

4.4.5 Overdiagnosis rate

For definition Gleason 3+3, the rate of overdiagnosis for T2W alone was 20.3% (95% CI 15.6-26), for T2+DWI was 22.4% (95% CI 17.4-28.2) and for T2+DWI+DCE was 24.4% (95% CI 19.3-30.3). For the secondary definition of any cancer excluding UCL2, the rate for T2W alone was 26.4% (95% CI 21.1-32.5), for T2+DWI was 31.3% (95% CI 25.6-37.6) and for T2+DWI+DCE was 34.1% (95% CI 28.3-40.5). There was no significant difference for overdiagnosis between T2W, T2W+DWI and T2W+DWI+DCE, respectively ($p = 0.274$, $p = 0.077$, $p = 0.564$).

4.4.6 Interobserver Variability

The weighted agreement on the double-read of the full mpMRI sequence was 87.0% (K=0.52, SE=0.10) indicating good agreement. When the mpMRI scores for each reporter were compared with TTPM biopsy histology, there were minimal differences between each reporter in terms of AUROC analyses (reporter 1: AUROC 0.76 [95% CI 0.63–0.90] vs reporter 2: AUROC 0.75 [95% CI 0.61–0.89]).

4.5 Discussion

4.5.1 Principle findings

This chapter has presented data from the PICTURE trial showing that the addition of DWI reduced the number of indeterminate lesions compared to T2 alone. The addition of DWI and DCE was marginal in terms of diagnostic accuracy at Likert ≥ 3. There was no difference in the rate of overdiagnosis between each sequence and good overall agreement between reports. This suggests that an abbreviated protocol including T2 and DWI could function as an effective screening test due to high diagnostic accuracy, low rate of indeterminate lesions and low overdiagnosis rate.

Given the safety concerns associated with DCE and lack of evidence for diagnostic superiority, these findings suggest that it will be necessary to omit the DCE sequence from an MRI screening protocol. The importance of DCE has already been reduced in PI-RADSv2 to a classification role for equivocal lesions in the peripheral zone[214] with the PI-RADS committee recognising that elimination of DCE may be a logical step once sufficient high-quality evidence is available[215].

There is reason to be cautious in moving towards a fast MRI screening protocol which includes only T2W. Although the results shown in Figure 8 and Figure 9 suggest that DWI and DCE did not improve the AUROC values, there was a shift in the distribution of Likert scores. The addition of DWI had a useful role in upgrading Likert 3 lesions and improving the reporter's scoring confidence for other lesions. The addition of DWI and DCE led to a reduction in Likert 3 lesions from 45% to 33% and a corresponding increase in Likert 5 lesions from 12% to 31% (T2W+DWI) or 38% (T2+DWI+DCE). This shift has also been observed in similar studies using PI-RADSv2 where DCE has a role distinguishing equivocal lesions in the PZ[216].

4.5.2 Comparison with previous studies

This chapter provides further support to several meta-analyses which have reported that the incremental benefit from DCE is marginal or non-existent[217, 218]. The largest meta-analysis to date reported no significant difference in sensitivity (mpMRI: 86%, 95% confidence interval [CI] 81-90; bpMRI: 90%, 95% CI 83-94) or specificity (mpMRI: 73%, 95% CI 64-81; bpMRI: 70%, 95% CI 42-83) with the summary AUROCs being comparable for mpMRI (0.87) and bpMRI (0.90)[219].

Subsequent to this, van der Leest et al [204] have reported that sensitivity for high-grade prostate cancer for both bpMRI and mpMRI was 95% (95% CI 91-97%) with specificity 69% (95% CI 64-73%). In this study, biopsy could be avoided in 49% for the bpMRI and mpMRI protocols. The recently published Danish BIDOC paired cohort clinical utility study in over one thousand men has also shown that bpMRI has good performance characteristics but was unable to compare their findings to mpMRI [123]. A recent meta-analysis, Bass et al[220], found the pooled sensitivity was 84% (95% CI 80–88%) and specificity 75% (95% CI, 68–81) for bpMRI.

In comparison to previous studies, a key strength of this chapter was evaluating the diagnostic performance of each sequence against transperineal prostate template mapping (TPM) biopsy. TPM biopsy provides an accurate reference standard for clinically significant prostate cancer due to the fixed 5mm sampling of the entire prostate. This approach minimises the methodological limitations from alternative reference tests such as transrectal ultrasound-guided (TRUS) 10-12 core systematic non-MRI guided biopsy which have an inherent random error, and whole-mount radical prostatectomy specimens which have an inherent selection bias.

4.5.3 Implications of findings

The present chapter serves to highlight that the DCE sequence might be safely omitted for screening without significantly impacting diagnostic accuracy. In this study, when DCE was omitted the NPV for clinically significant disease remained high for T2W (91% [79-98]) and for T2W+DWI (87% [70-96]). This suggests that a fast MRI protocol could still function as a triage test to rule out clinically significant prostate cancer in men. The addition of DCE identified a single extra case of clinically significant disease and no additional cases of significant disease were identified with the addition of DWI.

An effective screening test must offer a suitable balance between sensitivity and specificity. Given our findings, a biparametric MRI (bpMRI) including T2W and DWI-alone may be a reasonable approach to balance the reduction in indeterminate lesions with diagnostic accuracy for a screening test. These encouraging performance characteristics reflect the improvements in quality of DWI from modern MRI machines.

Endeavours that shorten the sequences without significantly compromising the detection rates of clinically significant prostate cancer would make faster MRI more feasible as a screening test by improving cost efficiency and throughput. There is already a growing

interest in bpMRI protocols which omit the administration of gadolinium-based contrast agents. The removal of contrast has advantages in terms of patient acceptability as well as reduced scanning time. The safety of gadolinium-based contrast has been questioned following reports that it can form depositions within the dentate nucleus and globus pallidus. These new bpMRI protocols will need to be evaluated in a screening population as one cannot infer the performance of tests which have been evaluated in a different population due to the effects of spectrum bias. Once the health setting changes to a population with lower prevalence of disease, the test will typically have a lower sensitivity and higher specificity.

Rather surprisingly, our findings suggest that it may be feasible using T2W scans alone on a 3T MRI to achieve a high level of diagnostic accuracy. A similar finding has been previously reported by Mertan et al[221] in a prospective study of 62 patients using a 3T MRI. In this study, T2W imaging alone had equivalent performance for cancer detection compared to the full combination of sequences as required in PI-RADSv2. However, these findings that a high diagnostic accuracy can be achieved using T2 should be treated with caution as the results are derived from a single high-volume centre using a 3T MRI scanner. A screening test must be a technique which can be replicated across readers with a range of experience which is not likely the case with T2w only.

In addition, there was a high number of equivocal lesions associated with the T2W alone protocol. It has been suggested that the higher proportion of equivocal lesions could be addressed by only performing the additional sequences in those cases where such a lesion is identified; however, this would require scans to be immediately reported or the patient to return for a second scan at a later date. For these reasons, I would not recommend that a T2W only approach is adopted at present without significant further research in multi-centre trials and across other patient populations.

Another implication highlighted by our results is that although the overall accuracy of mpMRI at Likert ≥ 3 was high across all combinations of sequences and definitions of clinically significant disease, there was a lower specificity. This lower specificity would make a Likert ≥ 3 threshold less attractive if persistent in a low prevalence population. In this respect a higher score might be needed in this population to avoid excessive unnecessary biopsies for likely benign MRI lesions. At Likert ≥ 4 there was a difference in diagnostic accuracy between the sequences, particularly T2w alone had a sensitivity between 51-65% depending on the definition of significant disease. In contrast sequences which included DWI and/or DCE had

sensitivity between 72-88%. Based on these findings in future chapters I will evaluate two thresholds to denote a screen-positive MRI; a score ≥ 3 and a score ≥ 4.

4.5.4 Limitations

This chapter has several limitations. First, these findings relate to an expert centre using a 3T MRI producing high-quality T2W images. It is acknowledged that this limits the reproducibility of my findings given that 1.5T is commonly used. In addition the finding that a high diagnostic accuracy can be achieved using T2 should be treated with caution due to the highly experienced reporter and state-of-the-art 3T scanner.

Second, this represents a whole-gland analysis and regional analysis differentiating PZ and TZ may highlight an added value of DWI and/or DCE. There is evidence that DCE has particular benefits in detecting the PZ lesions and differentiating equivocal from benign lesions[222]. Although there was a marginal improvement in diagnostic accuracy with the addition of the DWI and DCE, it did not reach statistical significance

Third, the study had a heterogeneous patient population who had a previous prostate biopsy. These findings will not relate to a screening population so the sequences validated in this study will be evaluated in a screening population in the subsequent chapters.

Fourth, there is evidence that DCE might improve tumour volume estimation[223] which might improve the targeting of biopsy. The effects on targeting was not evaluated in this study as TPM is a non-targeted procedure. In addition TPM is not a suitable reference standard for evaluation of extraprostatic extension so the impact of bp-MRI on staging could not be assessed in this study.

Last, and importantly, we cannot test the clinical utility of a bpMRI pathway compared to a mpMRI based pathway in which decisions on biopsy are made without the use of DCE in the bpMRI pathway. This is necessary since radiologists may score differently when they know that patients have not had the DCE sequences.

4.6 Conclusion

The present chapter serves to highlight that specific sequences might be safely omitted without impacting diagnostic accuracy. I have shown that a fast MRI protocol without DCE can achieve a high level of diagnostic accuracy using a 3T MRI in men with a prior biopsy. To minimise harms from equivocal lesions, the fast MRI protocol needs to include the DWI sequence.

Omitting this sequence led to a high rate of indeterminate (Score 3) MRI lesions which will present a clinical dilemma for a screening population. Given these findings, which are supported by other large studies extensively reviewed in the literature, that T2W + DWI offers the optimum balance between reduction in indeterminate lesions and diagnostic accuracy, this sequence will be utilised in subsequent chapters as the basis for the protocol for a fast MRI.

Chapter 5 – Design and recruitment into a population based screening trial: The IP1-PROSTAGRAM study

5.1 Overview

In order to evaluate the performance of fast MRI I designed in collaboration with my supervisors a prospective, population-based, blinded, screening study called the IP1-PROSTAGRAM trial. Previous population-based screening studies have had low screening uptake among certain ethnicities and socioeconomic groups. In IP1-PROSTAGRAM, a wide range of recruitment strategies were evaluated including a targeted recruitment strategy which aimed to increase participation among these harder-to-reach groups. This chapter evaluates the outcomes of these strategies and includes content which has been published in JAMA Onc[v].

5.2 Introduction

A key issue for population based screening studies is ensuring that there is a diverse uptake across different ethnicities and levels of socioeconomic status. It is recognised that there are differences in prostate cancer incidence and mortality rates between men of African, Asian and European ancestry. Evidence from the UK[224], USA[225], Africa[51] and the Caribbean[226] has demonstrated that the risk of prostate cancer is significantly higher in black men compared to white men. Additional epidemiological data provide further support when age standardised mortality rates are compared across world regions (Figure 10).

The relationship between prostate cancer mortality and ethnicity has been complicated by the interaction with socioeconomic disparities. Socioeconomic status is a major determinant of mortality among other tumour types with established screening programmes such as colorectal, breast and cervical cancer[227-229]. In prostate cancer a similar relationship has been shown, driven by lack of access to and use of healthcare services among lower socioeconomic groups[230]. Previous studies found that men within a lower socioeconomic group have a two-fold increased risk of dying from the disease compared to those from a higher socioeconomic groups (HR 2.0 95% CI 1.5-2.6)[231].

[v] Eldred-Evans, D., et al. (2021). "Population-Based Prostate Cancer Screening With Magnetic Resonance or Ultrasonography: The IP1-PROSTAGRAM study." JAMA Oncology (in press)

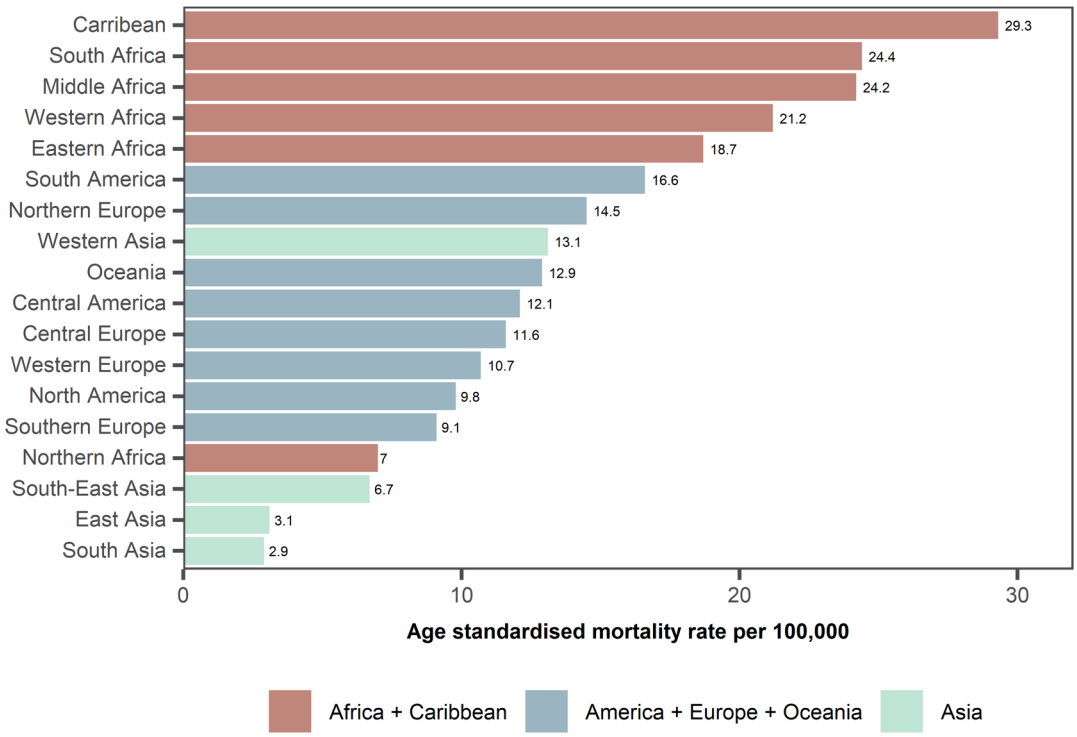


Figure 10: Prostate Cancer Mortality Rates by geographic area. Data from Global Cancer Statistics, 2012[232]. A clear epidemiological trend is shown with men from African and Caribbean countries (red) having a high mortality rate compared to other countries (blue). The lowest level of mortality is seen in regions of Asia (green).

Due to these disparities in mortality, it is essential that population-based screening studies have a recruitment strategy which can successfully engage a diverse population across different ethnicities and socioeconomic status. However, previous prostate cancer screening trials have not been able to recruit a diverse study population. The percentage of black men in PLCO was 4%[62] and in CAP less than 2%[37] while ERSPC did not report any ethnicity statistics[233]. The participants in PLCO[234] and CAP[235] were also of a higher socioeconomic status compared to the general population. Due to this lack of diversity within the previous studies, it has been recommended that subsequent studies need to pay special attention towards recruiting non-white males and those from a lower socioeconomic status[236].

On this basis, a prospective sub-study was nested within the IP1-PROSTAGRAM trial to examine a range of recruitment strategies and their ability to recruit a diverse population. The aim was to address inequalities in uptake and to evaluate the optimal recruitment method(s) to use in any subsequent future study of fast MRI screening.

5.3 Methods

5.3.1 Study Design

IP1-PROSTAGRAM was a prospectively registered, population-based, paired screen-positive cohort study which recruited men aged 50-69 years in the UK. Figure 11 provides a overview of the study design and participant flow. This chapter reports the results of the 'recruitment' section of the flow chart and subsequent chapters will address other aspects of the trial.

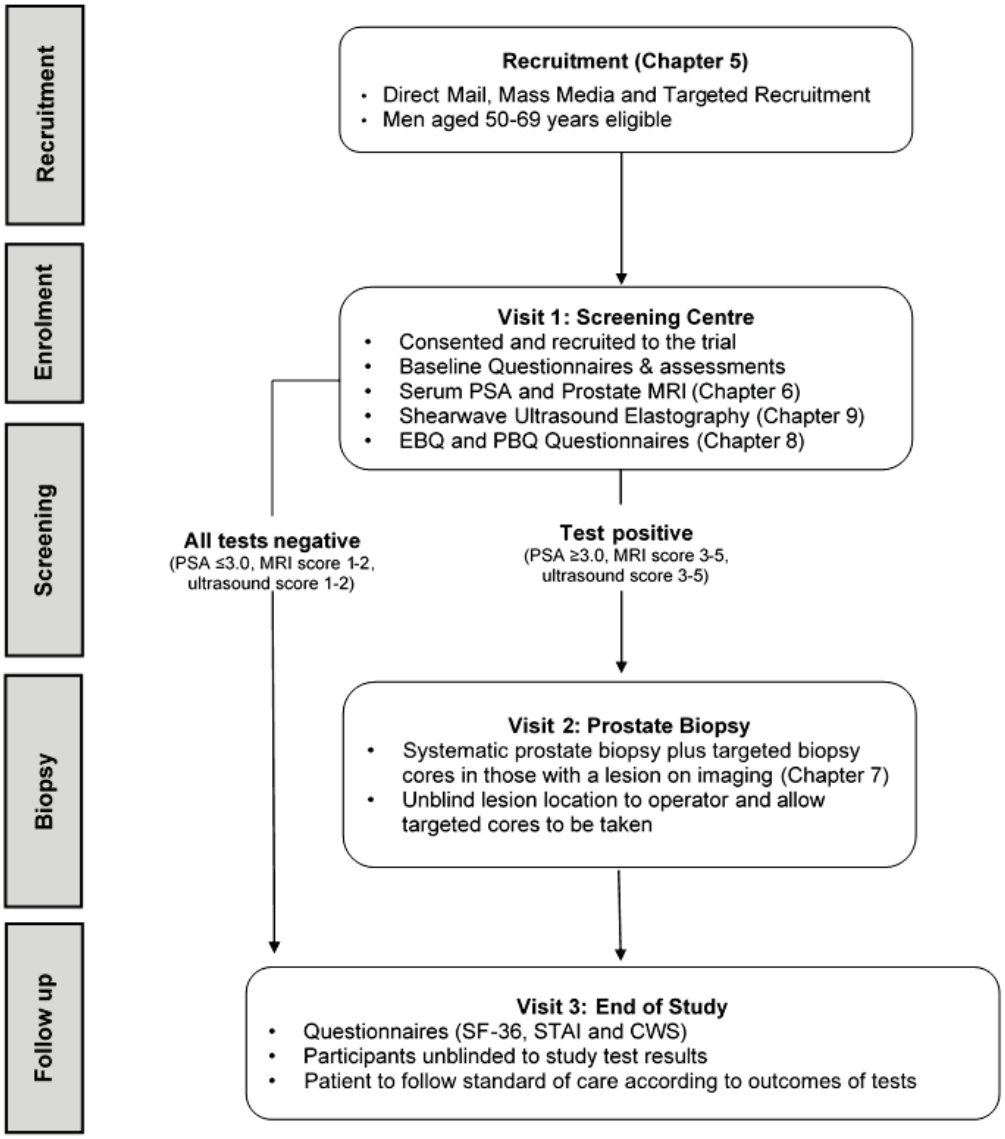


Figure 11: IP1-PROSTAGRAM Trial Study Schema.

In brief, IP1-PROSTAGRAM was designed to compare the performance of these new imaging techniques for prostate cancer screening. The primary outcome was to compare the positive test rate of a fast MRI with PSA and this is reported in Chapter 6. The study design incorporates multiple sub-studies which will be separately reported through Chapters 7-10.

Within the paired screen-positive design participants received a fast MRI, transrectal ultrasound and PSA test. A biopsy was performed if any screening test was suspicious for prostate cancer. All participants completed health quality of life (HRQoL) and patient reported experience (PREM) questionnaires before and after each screening test. These outcomes will be reported in Chapter 8.

The study was approved by the UK National Research Ethics Committee (8/LO/1338) and it was conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. All participants were provided with written informed consent. The trial was overseen by an Independent Trial Steering Committee and reported to the Standards for Reporting Diagnostic Accuracy[237]. I registered the IP1-PROSTAGRAM study with ClinicalTrials.gov (number NCT03702439) and ISRCTN (number 43502108).

5.3.2 Recruitment Strategy Design

The recruitment materials were designed to limit barriers to uptake, particularly amongst key ethnic minority and lower socioeconomic groups. Advice was taken from the UK Lung Cancer Screening Trial (UKCLS)[238] and the Department of Behavioural Science and Health, University College Hospital[239] who had designed similar campaigns to target lower socioeconomic groups for lung cancer screening.

Research from these groups had shown that excessive health related messaging in screening leaflets triggers a fearful response. This can elicit defensive, and therefore avoidant, behaviour and lower engagement from potential participants[240]. There is also evidence that keeping information brief and simple is key to increasing engagement, particularly when it comes to lower socioeconomic groups who may not be as health literate. A further barrier, faced by many screening tests but especially in prostate cancer, is that a large proportion of our respondents were likely to be asymptomatic, creating less of a sense of urgency around taking part.

Therefore, a recruitment strategy was developed to prioritise:

- Approachability: Reducing avoidant fear responses by framing the screening tests within the context of an overall ‘prostate health check’ and ensuring a down to earth ‘tonality’ through visuals and language

- Accessibility: Reducing the potential for disengagement by ensuring that we did not overwhelm respondents with information, using lay language and streamlining the amount of information given at different stages via a stepped approach

- Relevance: Clearly highlighting that this was for all men aged 50-69 with or without symptoms

5.3.3 Recruitment Methods

The recruitment methods for IP1-PROSTAGRAM were categorised into three broad categories:

1. Direct Mail Strategy: The use of mass mailing or SMS to invite participants registered with a general practitioner. This is a common form of recruiting to population-based screening studies and mass mailing via primary care has previously been used in the CAP and ERSPC screening studies.

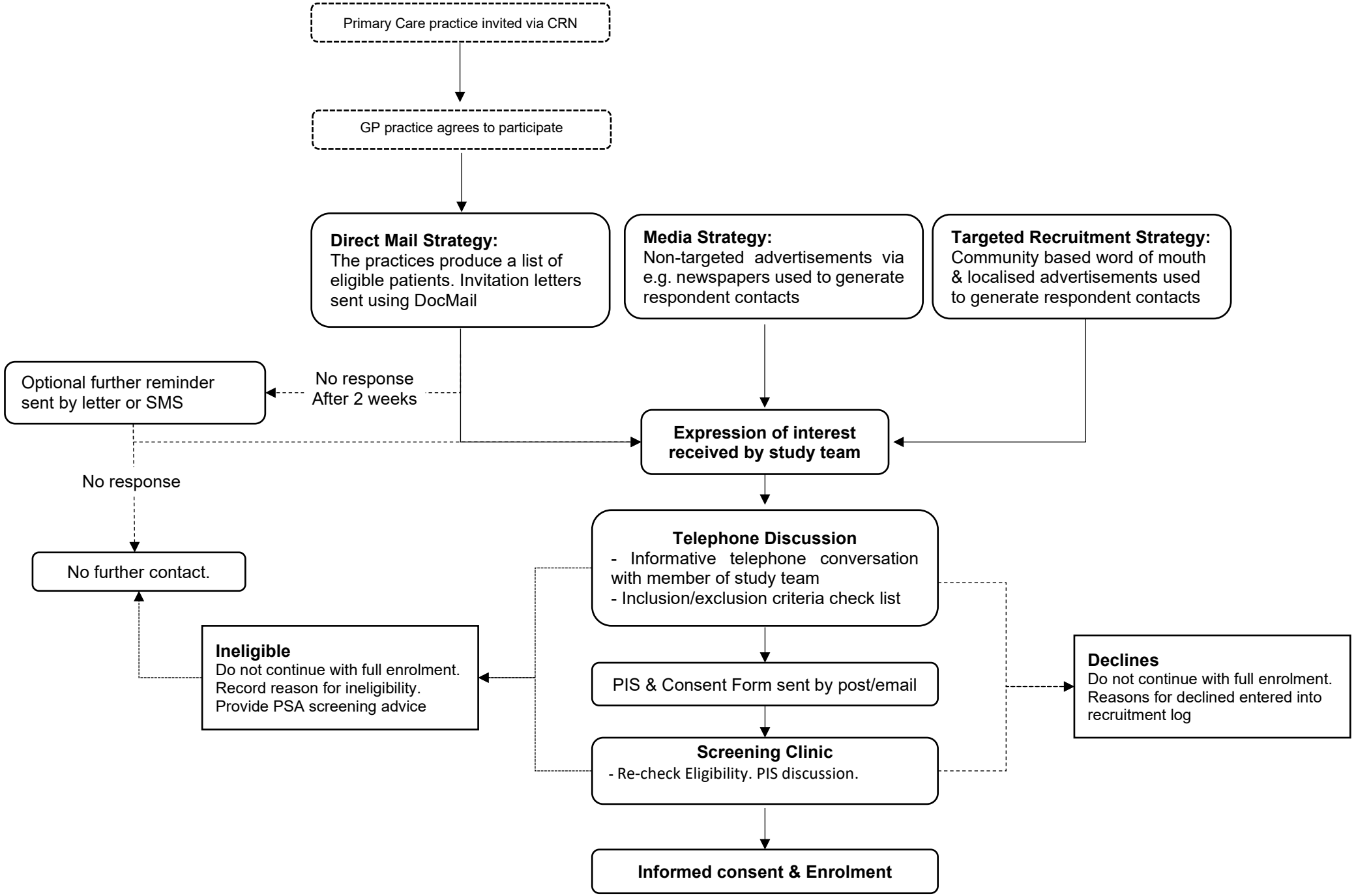
2. Media Strategy: This involved the use of print, broadcast or social media to inform men across the UK about the trial. It was not targeted towards any ethnic or socioeconomic group. A similar recruitment strategy was used for the PLCO study.

3. Targeted Recruitment Strategy: Targeting community hubs, and involving community group leaders as advocates of the study as well as general word-of-mouth recruitment.

All men who expressed an interest in response to any strategy were telephoned by the study team who confirmed eligibility criteria and booked them into clinic using a bespoke booking system designed for the IP1-PROSTAGRAM study. Further details are in Appendix III.

The following sections explain each recruitment method in detail and the complete recruitment flow chart is shown in Figure 12:

Figure 12: Recruitment flow chart



5.3.3.1 Direct Mail Strategy (Non-Targeted)

This recruitment method was performed by seven primary care practices in North West London. Primary care practices in the region were approached and those who expressed an interest ran database searches using pre-defined eligibility criteria. The filters for the searches excluded certain groups of men following criteria set out in Table 10:

Table 10: Primary care search criteria for excluding non-eligible men

1. A prostate-specific antigen level or prostate MRI in the last 2 years
2. An infection of the urinary tract or prostatic inflammatory disease in the last 6 months
3. A previous diagnosis of tumour of prostate or treatment for prostate cancer
4. Contraindications to PSA or MRI such as a needle phobia, claustrophobia, MRI incompatible devices, BMI 40kg/m2, glaucoma, low mobility, degenerative neurological disease or patients on home oxygen
5. Contraindications to prostate biopsy such as congenital bleeding disorders or anticoagulation
6. Co-morbidities which reduce life expectancy to <10 years such as metastatic cancer, inclusion on palliative care register, Acquired immunodeficiency syndrome, Congestive Heart failure, Chronic obstructive pulmonary disease (Medical Research Council (MRC) dyspnoea scale 4-5), myocardial infarction or unstable angina in last 12 months, Portal Hypertension/Liver cirrhosis, Chronic kidney disease 4 or 5

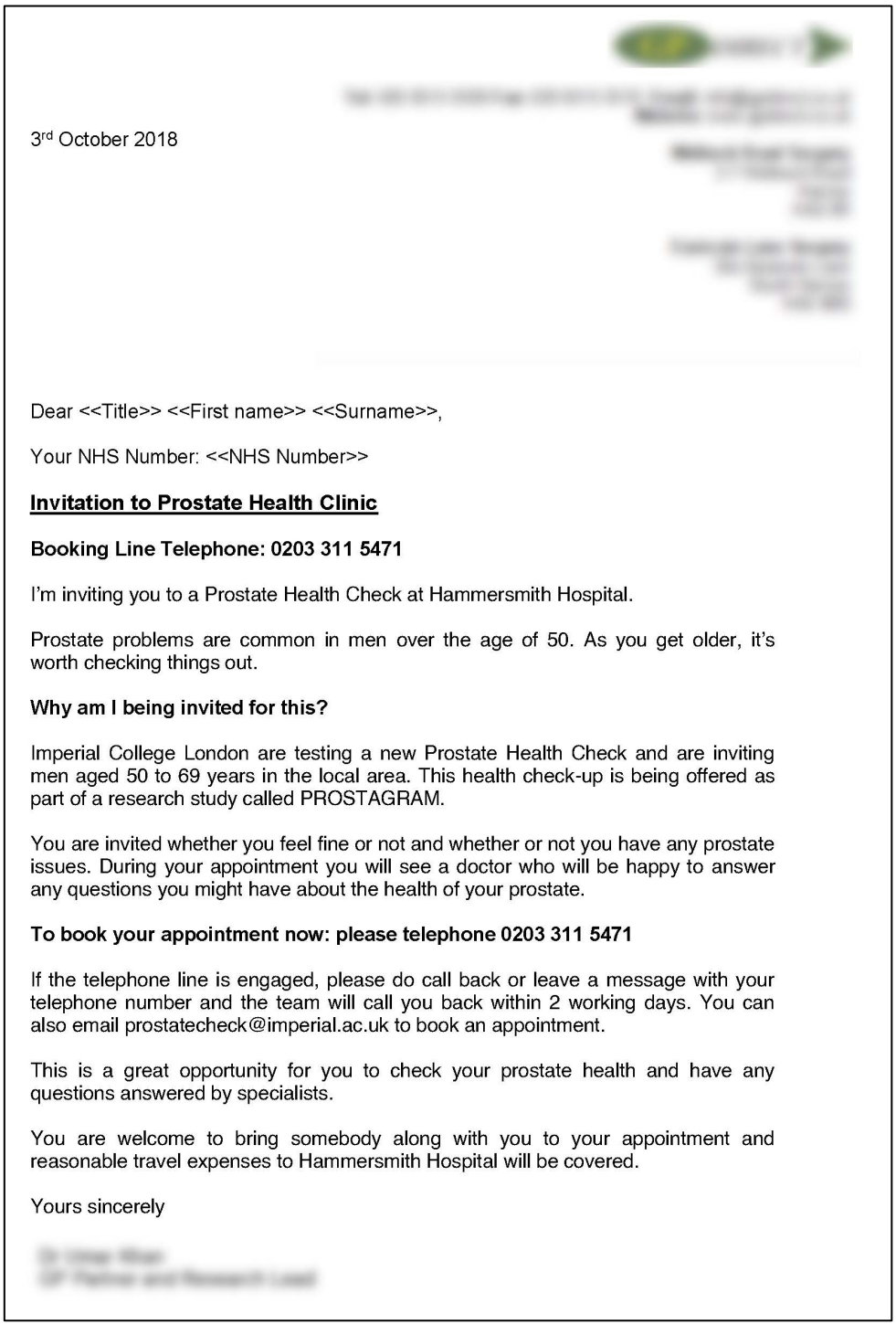
A general practitioner further screened lists to remove any men with other co-morbidities and/or frailty which would have meant that an individual's life expectancy would limit their benefit from screening or where there were other reasons why it may be inappropriate for the patient to receive an invitation. The process for letter invitation involved uploading the final patient list to an online mailing company (Docmail).

All men from the primary care list were sent an invitation letter (Figure 13) with an information leaflet (Figure 14) or an SMS (Figure 15) depending on the GP practice policy. The date of birth and postcode were provided by primary care practices for all men who were invited for recruitment purposes.

The letter was designed to reduce the key barriers identified above. Key to the strategy was reducing fear based barriers by inviting potential participants to a 'Prostate Health Check'

rather than a 'Prostate Cancer Screening Test'. The ability to bring someone along as support was also included and the letter was signed by the person's named GP for personalisation.

To aid approachability and accessibility simple language was used throughout. The letter also made clear that the health check was for men aged 50-69 whether they 'feel fine or not'. Clear calls to action were also included by prominently displaying the telephone number for booking an appointment at two points in the letter.



3rd October 2018

Dear <<Title>> <<First name>> <<Surname>>,

Your NHS Number: <<NHS Number>>

Invitation to Prostate Health Clinic

Booking Line Telephone: 0203 311 5471

I'm inviting you to a Prostate Health Check at Hammersmith Hospital.

Prostate problems are common in men over the age of 50. As you get older, it's worth checking things out.

Why am I being invited for this?

Imperial College London are testing a new Prostate Health Check and are inviting men aged 50 to 69 years in the local area. This health check-up is being offered as part of a research study called PROSTAGRAM.

You are invited whether you feel fine or not and whether or not you have any prostate issues. During your appointment you will see a doctor who will be happy to answer any questions you might have about the health of your prostate.

To book your appointment now: please telephone 0203 311 5471

If the telephone line is engaged, please do call back or leave a message with your telephone number and the team will call you back within 2 working days. You can also email prostatecheck@imperial.ac.uk to book an appointment.

This is a great opportunity for you to check your prostate health and have any questions answered by specialists.

You are welcome to bring somebody along with you to your appointment and reasonable travel expenses to Hammersmith Hospital will be covered.

Yours sincerely

Figure 13: Design of direct mail from a Primary Care Practice .Details of the practice have been anonymised.

A leaflet accompanied the letter which expanded upon what men could expect. This enabled a stepped approach to the amount of information delivered, helping reduce the burden level. The design brief for the visual style of the leaflet was to create an approachable and distinctly non-clinical feel. The final leaflet design was the result of a patient-led design collaboration. This process involved a competition amongst graphic designers and over one hundred design options submitted. A shortlisting process was led by the patient representatives of the trial management group and the final design was chosen from the shortlist by the PPI group using an online voting system.

This final design combined a bright welcoming colour palate and rounded shapes. In line with previous research, the designer was briefed to create sections of easily digestible information with infographics and visuals to help navigate readers through the information. Similar to the letter, a clear call to action was included with the booking line telephone number presented prominently on both side of the leaflet. The leaflet continued the use of accessible language throughout and relevance was highlighted by flagging that the health check was for men who may not have noticed any problems.

A. Leaflet Front

B. Leaflet Back

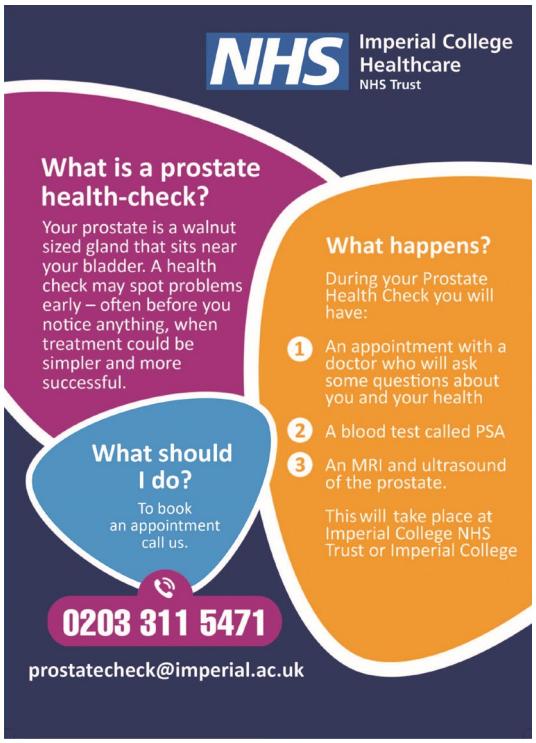


Figure 14: Information leaflet designed for the direct mail strategy.

Primary care practices were provided with an option of SMS messages which could be sent instead of direct mail depending on practice preference. This method was selected by two primary care practices and the SMS message was as follows:

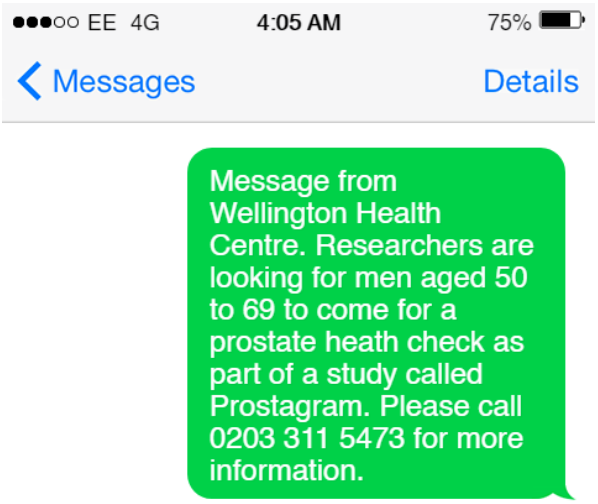


Figure 15: Example of SMS message sent via primary care practices.

5.3.3.2 Media Strategy (Non-Targeted)

Multiple non-targeted media strategies were employed concurrently during the study period. These included traditional advertisements via newspaper, radio and websites in combination with newer media strategies using social media channels:

1. Newspaper and radio advertisements: Adverts were placed in newspapers within the London area. The newspaper adverts were designed to be simple, clear and suitable for grayscale printing. Due to space constraints in newspapers, limited information could be provided on the adverts and they were designed to trigger potential participants to contact the study team for further information about the study. The newspaper adverts were supplemented by mentions of the study on radio stations covering the North West London area.

2. Social media: The social media campaign focused on platforms such as Twitter. Relevant accounts with high social capital in the context of prostate cancer posted information about the trial on their Twitter or Facebook feeds (Figure 16). These linked to the study website where potential participants could learn more about the trial.

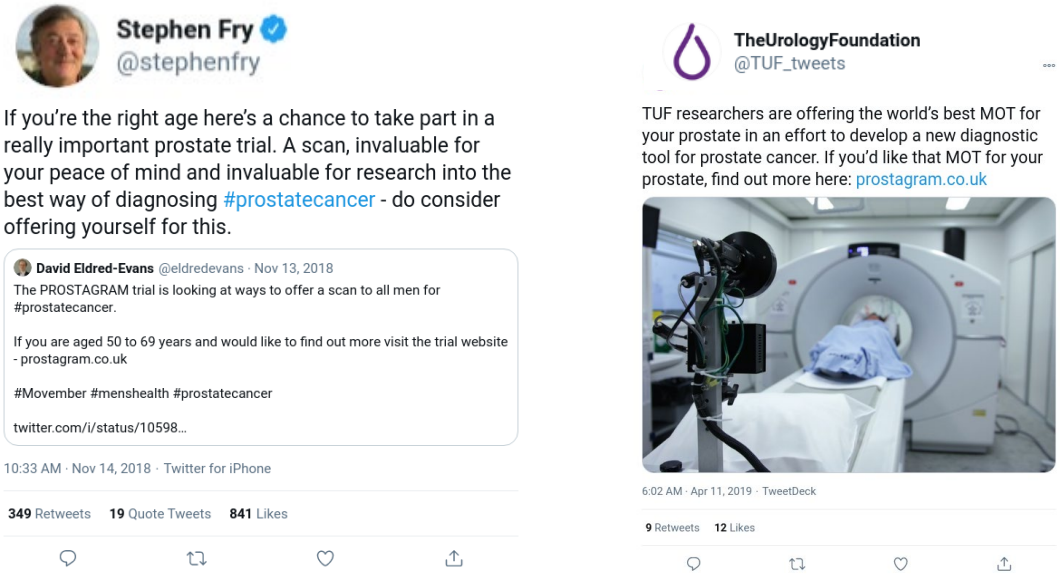


Figure 16: Examples of social media recruitment strategy on Twitter.

3. Study website: An official study website was set up where participants recruited via radio, newspaper or social media could directly register for the study (http://imperialprostate.org.uk/prostagram/) (Figure 17). This was supplemented by an additional website set up through the Imperial Clinical Trials Unit http://www.imperialclinicaltrialsunit.org/trials/prostate-cancer-screening-trial-using-a-group-of-radiological-approaches-including-mri-and-ultrasound/.

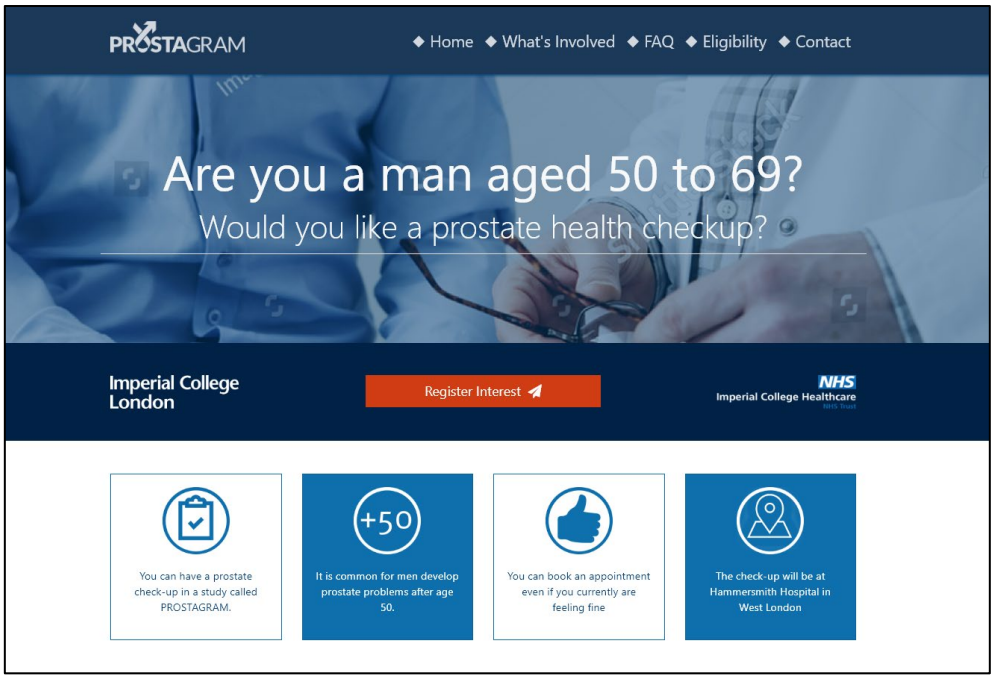


Figure 17: Study website (http://imperialprostate.org.uk/prostagram/). Further screen shots on following page.

5.3.3.3 Targeted Recruitment

The targeted recruitment strategy was designed with the specific aim of promoting the study amongst black men and lower socioeconomic groups. A key barrier to participation by ethnic minority groups in clinical trials is mistrust of the medical community and medical research[241]. We sought to tackle this in several ways:

1. Building an ethnically diverse recruitment team from the target population who were key to ensuring the messaging was culturally appropriate.

2. Team members were also empowered to become minority recruitment champions and use their community links where appropriate.

3. Community leaders were engaged as advocates for the trial via the recruitment team as well as faith and online communities. We worked in collaboration with them to drive word of mouth around the study, enhancing the depth of our reach with black communities.

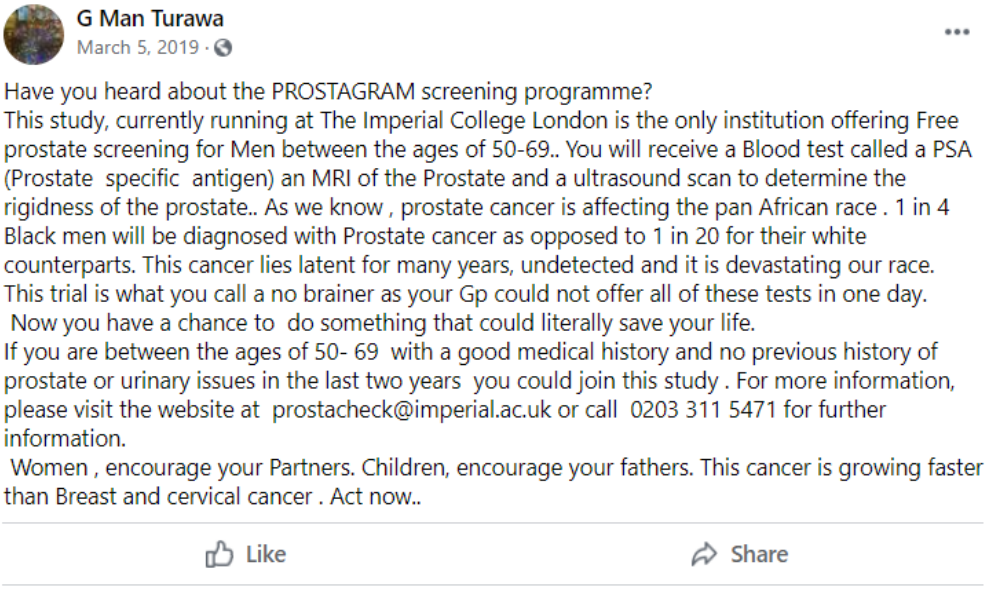


Figure 18: Example of word of mouth exchange within local community on Facebook regarding the IP1-PROSTAGRAM trial (shared with consent of participant).

The second part of the strategy involved a localised poster campaign in areas identified as having high levels of deprivation and ethnic diversity using the index of multiple deprivation (IMD2019). These areas were targeted with posters and leaflets in community hubs and gathering places such as church halls, libraries, supermarkets and pubs. The poster was developed through the same patient-led design process described above for the leaflet above. Given the context in which the poster would be viewed, key to the design was to

attract attention from the relevant target group and streamline the information. Therefore, a bold colour palate was used with a question led headline designed to trigger interest in all men aged 50-69.

Figure 19: Poster and leaflet targeted recruitment campaign.

Additionally, the poster design was focused around reducing the same barriers discussed in the previous sections. To reduce anxiety and avoidant behaviour there was no mention of cancer, and to aid accessibility there was a focus on clear and simple language and a sectioned information structure. Relevance was highlighted to those that don't feel they have problems, and a sense of urgency was generated through key phrases such as 'sooner rather than later' and clear calls to action with a prominent telephone number of the booking line.

5.3.4 Outcome Measures

5.3.4.1 Sociodemographic Measures

The outcome measures were based on ethnicity and the Index of Multiple Deprivation (IMD). Ethnicity was self reported by participants at the screening clinic according to a standardised list of official ethnic groups provided by the Office for National Statistics. Participants reported as 'mixed' or 'other' for ethnicity were grouped into a single category due to low numbers in each group.

The IMD was used to measure socioeconomic status and is a widely used measure of deprivation in England. It a composite measure combining seven weighted domains of deprivation for each area including Income; Employment; Education; Skills and Training; Health and Disability; Crime; Barriers to Housing Services; Living Environment.

The IMD can be presented as a rank from the most deprived area (1) to the least deprived area (32844). For categorical presentation these can be divided into quintiles and quintile 1 is equivalent to areas 1 to 6,569 (most deprived) increasing to quintile 5, equivalent to areas 26,275 to 32,844 (least deprived).

5.3.4.2 Primary and Secondary outcomes

The primary outcome compared ethnicity and IMD of the study participants with the local population and between recruitment strategies. Secondary outcomes included comparing a range of sociodemographic variables across each recruitment methods. A cost comparison of each recruitment method was undertaken and a comparison between responders and non-responders in terms of age and IMD was completed.

5.3.5 Statistical Analysis

All analyses were conducted in R Version 4.0.2[213] using R Studio Version 1.3.1073. Chi-squared squared tests of independence were used to compare differences in ethnicity and index of multiple deprivation between each screening tests. Other differences in sociodemographic variables such as education levels, marital status qualifications, BMI, family history of prostate cancer, CCI and smoking history were compared with chi squared for categorical or Kruskall Wallis for continuous.

The monthly accrual rate for each targeted strategy was plotted along with overall study recruitment compared to the actual recruitment rates. The cost of each recruitment strategy was calculated where possible by dividing the total cost of the strategy by the number of men who responded. Costs were calculated on a per respondent basis rather than number recruited as the study was over-subscribed so respondents from certain strategies were reduced.

For the secondary outcome comparing responders and non-responders invited by letter the association between age, IMD and response to invitation were evaluated using binomial logistic regression. Adjusted odds ratios were used to compare the proportion of responders and non-responders by response rate and recruitment rate. T tests were used to compare the mean deprivation score and age between responders and non-responders.

5.4 Results

5.4.1 Recruitment Overview

Between September 20, 2018 and May 15, 2019, 1,316 expressions of interest were received from men in response to the recruitment strategies. A total of 387 expressions of interest were received due to the targeted recruitment strategy and 612 due to the media campaign. The direct mail strategy generated 317 expressions of interest from 1707 invitations letters.

A time series illustrating the numbers of expression of interest received during the study period by each recruitment strategy is shown in Figure 20. The direct mail strategy commenced in September 2018 and achieved a consistent rate throughout the study period. In contrast, the responses received due to the media and targeted strategies was intermittent and related to timing of study team interventions. The media campaign had minimal response until a tweet by a prominent prostate cancer patient with 12.6 million followers. This tweet had 9,396 views and generated 1,534 referrals to the study website. It led to 587 expressing an interest in the study over a 48 hour period.

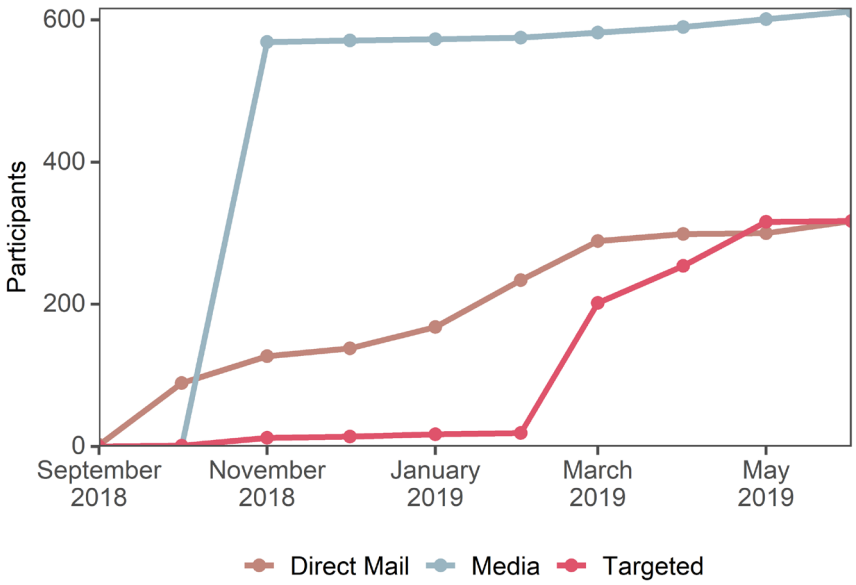


Figure 20: Cumulative expressions of interest received by each screening recruitment method.

From the 1,316 expressions of interest, 42.2% (n=554) of potential participants were contactable for telephone pre-screening. This included 312 from the direct mail group, 79 from the media campaign and 163 from targeted recruitment. From this group, 143 men were not booked into the screening clinic due to ineligibility (n=105), declining to participate (n=23) or not attending on the day of clinic (n=15). In total, 411 men attended the clinic and

were recruited into the study. This chapter compares the sociodemographic variables of these 411 men who were recruited into the study.

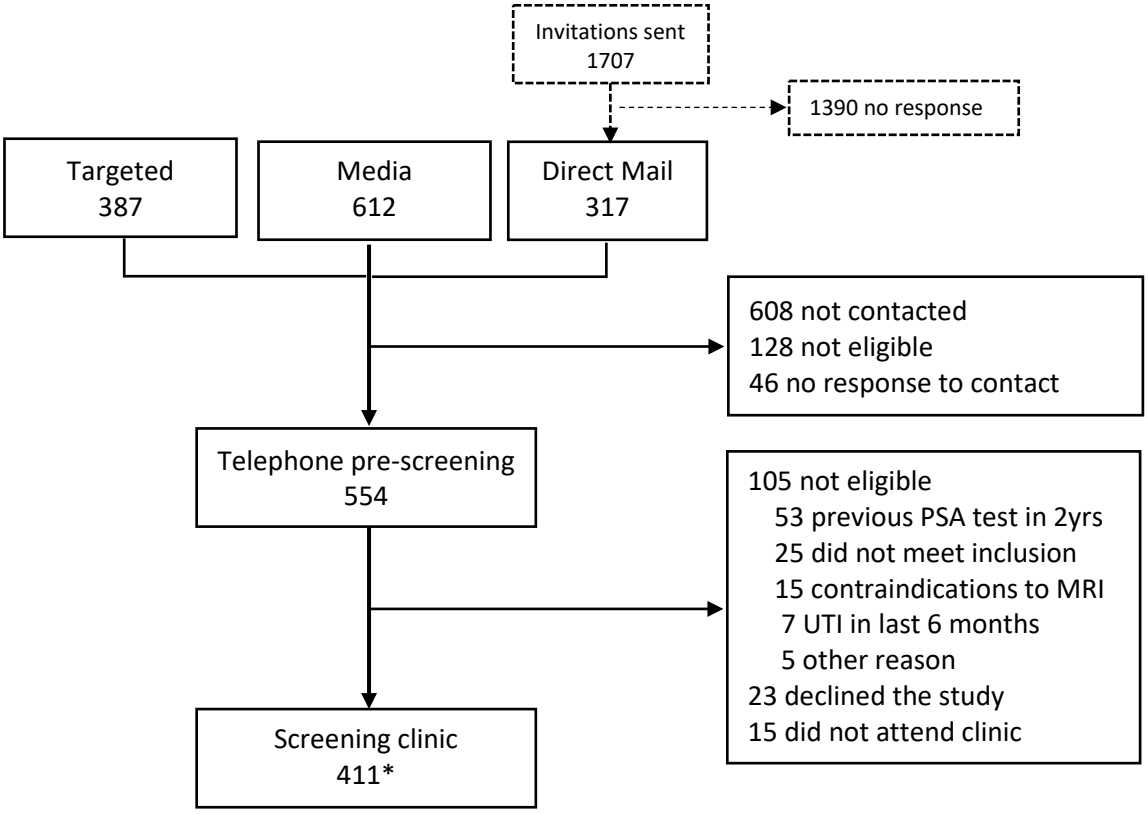


Figure 21: Expressions of interest received by each recruitment method and flow of participants.

The high number of expressions of interest meant that the study achieved rapid recruitment and it was completed 19 months ahead of schedule.

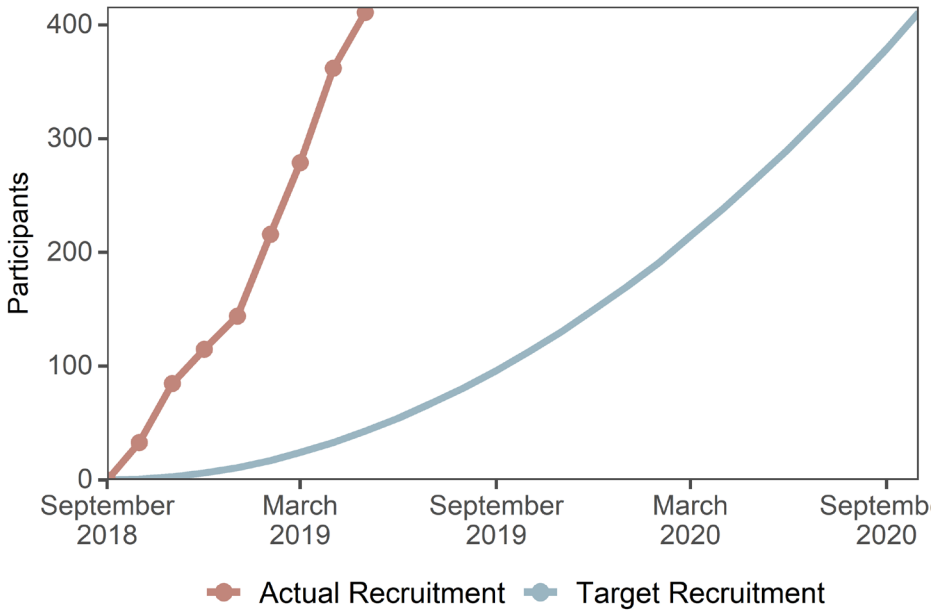


Figure 22: Cumulative total study recruitment compared to expected recruitment.

5.4.2 Primary Outcomes

5.4.2.1 Study participants compared to local population

Ethnicity: The ethnicity of 411 participants was distributed across White (38.0%), Black (32.4%), Asian (23.0%) and Other/Mixed (4.4%) ethnic groups. This can be compared to the ethnic distribution within the boroughs of West London where the study was predominately recruited. In total, the recruitment strategy recruited a higher proportion of black men than would be expected given the local population (Figure 23). This excess recruitment of black men was driven by the targeted recruitment strategy as discussed in the following sub-section.

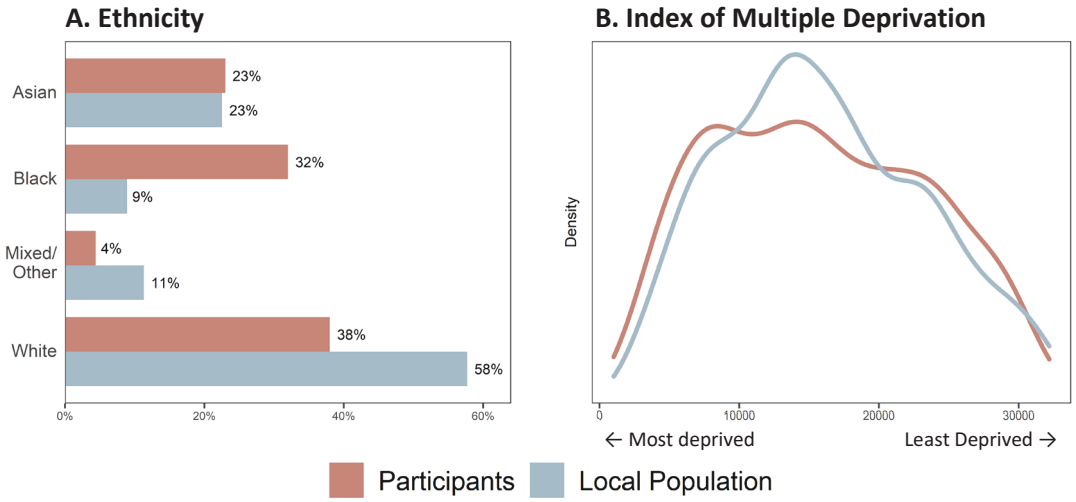


Figure 23: Ethnic group and Index of Multiple Deprivation in study participants compared to local population. Local population calculated from ethnic groups within boroughs of Chelsea, Hammersmith and Fulham, Harrow, Hillington, Islington and Kensington. Data from Office of National Statistics, Annual Population Survey 2018. (A) Bar charts of ethnicity by four major ethnic groups (B) Frequency Curve of Index of Multiple Deprivation.

Index of Multiple Deprivation : The comparison of IMD is shown in a Frequency Density plot in Figure 23. This suggests that the IMD distribution of study participants was similar to the local population. The distribution of participants is marginally left skewed towards recruitment of more deprived men. This is reflected when the IMD rank is presented as quintiles with the proportion of men increasing from 26% to 40% from least to most deprived IMD quintiles (Quintiles 4 and 5 vs Quintiles 1 & 2).

5.4.2.2 Ethnicity and Index of Multiple Deprivation by Recruitment Method

Ethnicity: A comparison of recruitment methods showed marked differences between the ethnicities recruited ($p<0.001$). The proportion of black men recruited by direct mail (8%) was similar to the prevalence of black men in the local population (9%). The number of black men

recruited by targeted recruitment was high (88%, n=115) and low for media recruitment (1.7%, n=1).

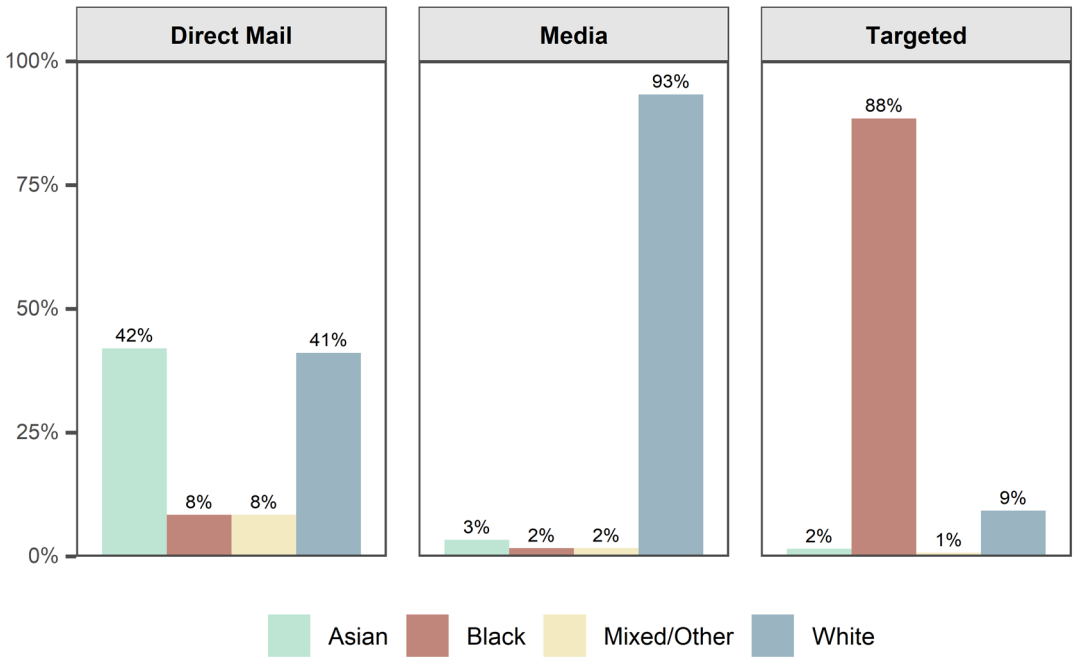


Figure 24: Ethnic groups recruited across each recruitment method.

Index of Multiple Deprivation: Each recruitment method also produced differences in IMD quintiles although not as marked as ethnicity. Direct mail recruited a close-to normal distribution, media recruited from the least deprived areas and targeted recruitment trended towards recruiting from most deprived areas.

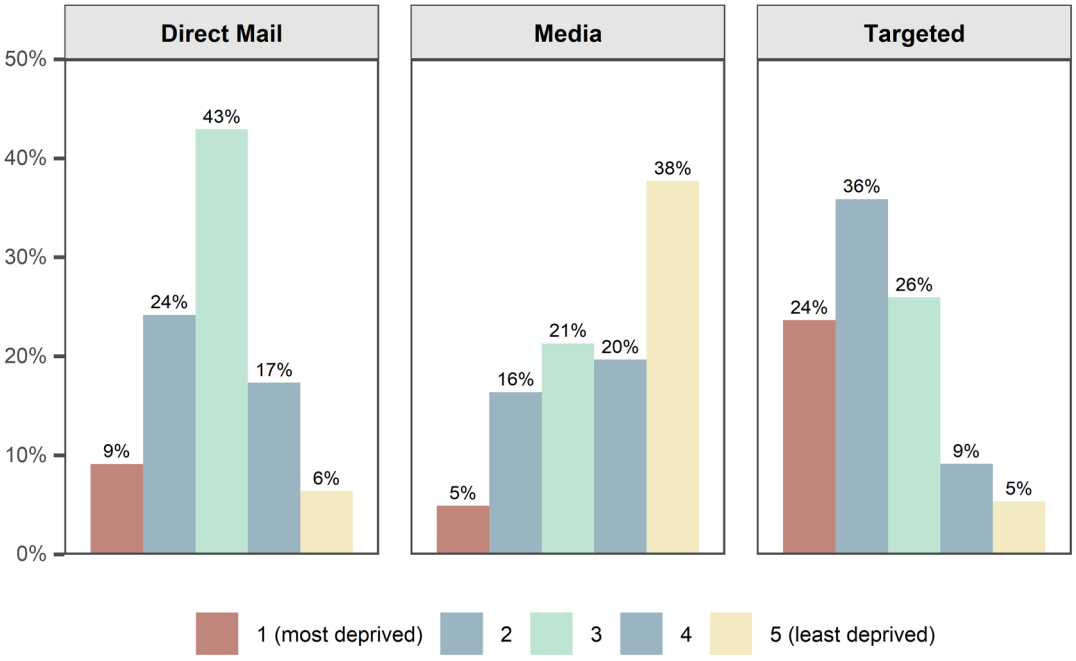


Figure 25: Index of Multiple Deprivation Quintiles across each recruitment method.

5.4.3 Secondary Outcomes

5.4.3.1 Other Sociodemographic Variables

Additional sociodemographic variables between recruitment methods are shown in Table 11. The targeted strategy recruited a younger cohort of men compared to media or direct mail (p<0.001). Men were more likely to report a significant family history of prostate cancer if recruited by the direct mail strategy. There was no significant difference in the level of qualifications, marital or employment status, BMI, co-morbidities or smoking history.

Table 11: Sociodemographic, Prostate Risk Factors and Medical History

	Direct Mail N = 219	**Media** N = 61	**Targeted Recruitment** N = 131	**p-value**[1]
Sociodemographic				
Age at invitation (yr)	58 (54-63)	58 (52-61)	55 (53-58)	<0.001
Married/Civil Partnership	173 (79%)	51 (84%)	96 (74%)	0.382
Employed	172 (80%)	42 (69%)	99 (76%)	0.254
Qualification				0.361
No Qualifications	20 (9.3%)	5 (8.3%)	5 (3.8%)	
GCSEs or O levels	35 (16%)	6 (10%)	28 (22%)	
A-levels o equivalent	24 (11%)	13 (22%)	11 (8.5%)	
University degree	119 (55%)	35 (58%)	76 (58%)	
Other	17 (7.9%)	1 (1.7%)	10 (7.7%)	
Prostate Cancer Risk				
BMI	27.2 (3.8)	27.0 (3.3)	27.8 (4.2)	0.3
Family History (1st)[2]	42 (19%)	0 (0%)	2 (1.6%)	<0.001
Family History (Any)[3]	53 (24%)	0 (0%)	28 (22%)	<0.001
Medical History				
IPSS Score	4 (2-8)	4 (2-9)	5 (2-10)	0.4
Number of Comorbidities (CCI)				0.2
0	165 (78%)	49 (80%)	113 (88%)	
1	40 (19%)	11 (18%)	12 (9.3%)	
2	6 (2.8%)	1 (1.6%)	4 (3.1%)	
Smoker				0.6
Current Smoker	28 (13%)	4 (7.1%)	16 (13%)	
Ex-Smoker	58 (27%)	20 (36%)	35 (28%)	
Never Smoker	126 (59%)	32 (57%)	72 (59%)	

[1] Statistical tests performed: Kruskal-Wallis test; chi-square test of independence;
[2] A first degree family member [3] A first or second degree family member

5.4.3.2 Variation in response to invitation by letter or text message

The seven Primary Care Practices were predominately located in more deprived areas as defined by IMD Quintile at the practice address (Table 12). Of 1,707 men who received an invitation via their Primary Care Practice, 18.6% (n=317) contacted the study team to express an interest and 81.4% (n=1,390) did not respond. A total of 219 men were recruited following further explanation and eligibility checks representing a recruitment rate of 12.8%. The remaining men either declined to participate (2.3% n=39), could not be contacted back by the study team (1.3%, n = 22) or did not meet eligibility criteria (2.9%, n=50). The most common reasons for not being eligible were a previous PSA within two years (n=22), insufficient English for consent (n=17) or a contraindication to MRI (n=5).

The invitations were sent by the primary care practice either by letter (80.2%, n=1,370) or text message (19.7%, n=337). The response rate from the letter was significantly higher than from the text messages (22.7%, 95% CI 20.5-25.0 vs. 5.6% 95% CI 3.4-8.7, p<0.001). The response rate between practices sending letters ranged from 13.8% to 28.0% while the rate for text messages was similar (5.5% to 5.8%). Table 12 provides a breakdown of the response rates by invitation method across each primary care practice.

Table 12: Response rates for letters and text messages by Primary Care Practice

	IMD* Quintile	Invitations sent	Response Rate**	Ineligible Rate^	Decline rate¥	Recruitment Rate^^
Letters						
Practice 1	4	500	28.0%	3.6%	3.2%	21.2%
Practice 2	1	253	26.9%	4.3%	9.5%	13.0%
Practice 3	4	222	14.9%	3.2%	1.8%	9.9%
Practice 4	4	235	20.4%	3.4%	4.3%	12.8%
Practice 5	3	160	13.8%	1.3%	3.1%	9.4%
Overall	3	1370	22.7%	3.1%	3.2%	13.8%
Text Messages						
Practice 6	2	200	5.5%	1.0%	4.3%	4.0%
Practice 7	1	137	5.8%	1.5%	0.5%	3.6%
Overall	1	337	5.6%	1.2%	0.7%	3.9%

All percentages calculated using invitation sent as denominator
* IMD by Lower layer Super Output Areas for each Primary Care Practice by postcode
** Proportion of expressions of interest received via telephone due to invitation
^ Proportion who found to be ineligible during telephone screening
¥ Proportion who declined or could not be contacted after expressing an interest
^^ Proportion who were recruited to the study

5.4.3.3 Sociodemographic variation in response to invitation letter

The comparison between responders and non-responders included invitations sent by letter. Details of the trial were sent by letter to 1370 men at participating primary care practices. A total of 22.7% (n=311) men responded to the invitation by letter and 77.2% (n=1059) did not respond. The frequency distribution of age and IMD scores is shown in

Figure 26. Responders were older compared to non responders (mean age 58.9 years (SD 5.36) versus 57.2 years (SD 5.27), p<0.001). In contrast, the distribution of IMD deprivation of scores is comparable and suggests that level of deprivation had no impact on uptake in this cohort. The mean IMD of non-responders was similar to responders (mean IDM 16,580 (SD 6371) versus 17,006 (SD 6972), p<0.001).

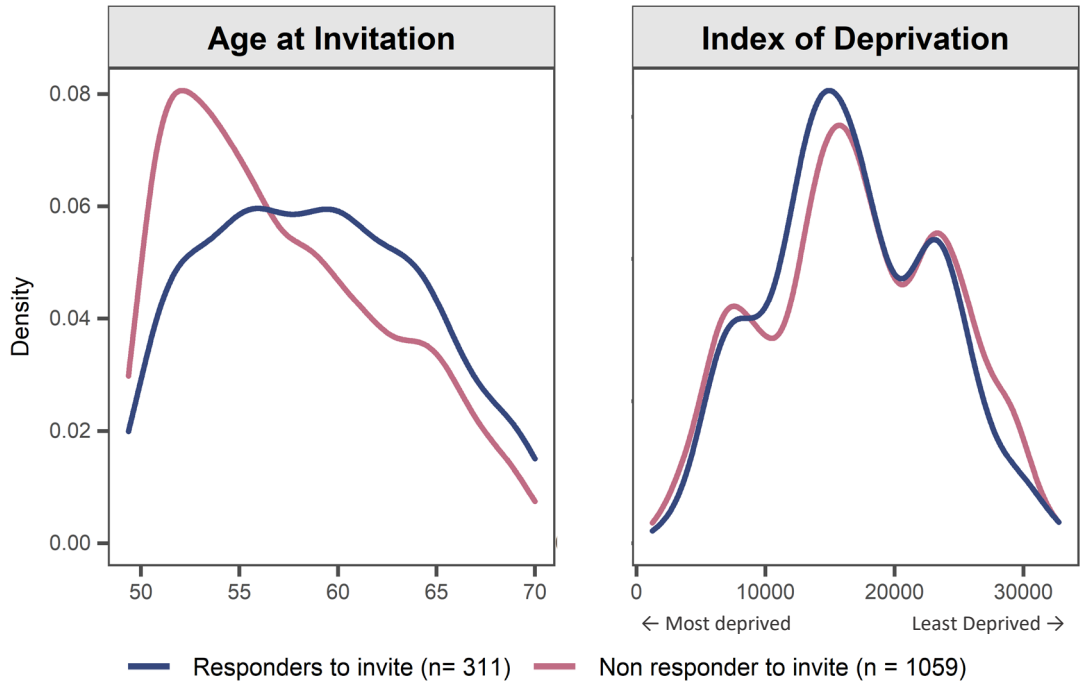


Figure 26: Frequency distribution of age and IMD rank for those responding (blue line) compared to those not responding to the invite (red line).

This trend is confirmed with multivariate analysis (Table 13) where there was a graded association across the four age groups, with lower response rates for men aged 50-54 years (reference category). Older age was associated with a higher response rate to screening and this trend was highly significant in 60 years and older group. There was no statistically significant difference in response rate by IMD rank.

In terms of recruitment rate, participants who agreed to be recruited were more likely to be older (65-69, OR 2.21 [95% CI 1.44-3.36], p<0.01). Similar to the response rate this gradient

had a highly significant trend. No association between IMD and decision to agree to be recruited to the study was found.

Table 13: Multivariate analysis for predictors of response and recruitment rate

Variable	Response to invitation*		Recruited^	
	Odds Ratio (95% CI)	p-value	Odds Ratio (95% CI)	p-value
Age (yr)				
50-54	1.00	Ref	1.00	Ref
55-60	1.62 (1.16-2.28)	0.005	1.81 (1.16-2.82)	0.009
60-64	2.17 (1.52-3.09)	<0.001	2.04 (1.40-1.70)	0.003
65-69	2.21 (1.44-3.36)	<0.001	2.21 (1.44-3.36)	<0.001
Index of multiple variations				
Quintile 1 (least deprived)	1.00	Ref	1.00	Ref
Quintile 2	1.47 (0.25-28.1)	0.72	1.89 (0.36-28.9)	0.76
Quintile 3	1.97 (0.44-37.4)	0.53	2.04 (0.48-37.8)	0.62
Quintile 4	1.49 (0.26-28.3)	0.71	1.43 (0.21-27.9)	0.81
Quintile 5 (most deprived)	1.42 (0.24-27.0)	0.75	1.22 (0.14-27.1)	0.82

For each variable, the odds ratio describe the odds of the outcome of the given category relative to the reference category.
* Analysis categorised individuals into non-responders or responders to the invitation by letter ^ Analysis categorised men into those recruited and not recruited (including non responders)

5.4.3.4 Recruitment yield and associated costs by each screening method

Data on total cost of each recruitment strategy and cost per respondent was available for invitations via letter, text messages, newspaper adverts and the local poster campaign (Table 14). Other aspects of the media and targeted recruitment did not accrue direct costs and are excluded from this analysis.

In terms of direct costs, there was a benefit from the direct mail strategy with cost per respondent calculated as £4.85 for letters and £3.94 for text messages. The cost per letter and leaflet was £1.12 compared to £0.22 for a text message. Therefore, although the response rate was lower for text messages, on a cost per response basis this was the most

cost efficient method of recruitment. The cost per participant for the newspaper adverts and local poster campaign were high at £110 and £14.54 respectively.

Table 14: Costs associated with each recruitment method*

	No. of Invitations	No. of responses	Total cost of strategy	Cost per response
Direct Mail Strategy				
Letters	1370	317	£1,539.80	£4.85
Text Messages	337	19	£74.88	£3.94
Media Strategy				
Newspaper adverts	-	6	£660[1]	£110
Targeted Recruitment				
Local poster campaign		10	£145.40[2]	£14.54

* Social media (media recruitment) and non-poster aspects of targeted recruitment excluded from cost estimates

[1] A quarter page advert (16cm x 13.4cm) for 1 week

[2] Printing costs for leaflets and posters. Staffing costs to distribute not included

5.5 Discussion

5.5.1 Principle findings

A range of recruitment strategies were evaluated within the IP1-PROSTAGRAM trial including a targeted recruitment strategy tailored to improve engagement of high-risk, hard-to-reach groups. The results show that the targeted recruitment strategy was capable of recruiting more black men and men from a lower socioeconomic group compared to the direct mail or media strategy. The use of this strategy led to 387 predominately black men requesting participation in the study. The high level of response meant that not all these men could be recruited and led to the study rapidly achieving its primary recruitment target of 411 men so that it completed recruitment 19 months ahead of schedule.

The findings for the media strategy were mixed. Although it generated a high number of responses (n=612), it was driven by one particular social media post which received 9,396 views and generated 587 expressions of interest in 48 hours. Men recruited via the media strategy were predominately white (93%) and from the least deprived socioeconomic group

(Quintile 5). These demographics were in keeping with the followers of the social media account holder[i vi].

Other attempts to generate media interest using traditional forms of recruitment such as newspapers and radio advertising generated minimal responses and were consequently associated with a high cost per expression of interest. In addition, alternative attempts to drive recruitment using social media profiles of prostate cancer charities generated few responses.

The final recruitment strategy involved using direct mail where the response rate from letter invitation was 4-fold higher than from text messages. A comparison of responders and non-responders to the postal invitations showed that this strategy was capable of recruiting men from a diverse spectrum of socioeconomic backgrounds. A similar comparison of the ethnicities recruited by letter to the local population suggested that we received the expected response rate from black men given the demographics of the local area. These findings suggest the letter and leaflet design was capable of recruiting black men and could be considered as a targeted method to recruit high-risk individuals in future trials.

In terms of differences in other sociodemographic variables, there was evidence that the direct mail strategy might have appealed to a marginally older age group compared to the media or targeted recruitment strategy. This was found in both the comparison of responders and non-responders as well as comparison between recruitment strategies. In addition, men recruited via direct mail were also more likely to report a first degree relative with a significant family history of prostate cancer. Both of these factors would need to be considered in selecting recruitment strategies for future trials to minimise the risk of skewing the risk profile of participants towards older men with a significant family history. Beyond these factors, there were no differences in education level, employment, marital status, smoking history, BMI or other medical co-morbidities.

There was also no evidence that the methodology of recruitment to a 'prostate health check' led to an over-recruitment of men suffering from urinary symptoms. During the design of the recruitment strategy one element of PPI feedback was that the 'prostate health check' wording might lead to men self-selecting based on the presence of pre-existing urinary symptoms. We responded to this potential issue by re-emphasising in recruitment materials

[vi] As per personal correspondence with Twitter Account holder and his team. The exact demographic figures of these Twitter Followers remain confidential and cannot be disclosed in this thesis.

that the prostate health was designed to be available for men without symptoms. Therefore, the recruitment material includes multiple phrases such as *"You can book an appointment even if you feel fine and have no problems"* and *"You are invited whether you feel fine or not, and whether or not you have any prostate issues"*. The results provide reassurance that this phraseology appears to have been successful given that the IPSS score among participants was low (4-5 points) and this was consistent across all methods of recruitment.

5.5.2 Comparison with previous studies

These findings are consistent with experiences of non-prostate cancer screening trials in which targeted strategies can significantly increase enrolment of high-risk and hard-to-reach individuals[242, 243]. Given that cultural perceptions play a significant role in determining the willingness of minority populations to participate in clinical trials[243], dedicated recruitment strategies provide the optimal method to build trust and alleviate specific cultural barriers to participation. Similarly to previous studies, enlisting 'cultural insiders' and staff members who had the trust of the target population was a highly successful method for reaching and recruiting minority participants.

Compared to other population screening trials, IP1-PROSTAGRAM recruited a higher proportion of black men as well as men from a lower socioeconomic background. Table 15 illustrates by comparison the demographics of this study to previous screening trials where there was a lack of ethnic diversity in study participants.

Table 15: Ethnicity Distribution in population screening trials compared to world population

Author (Year)	Study	Ethnicity		
		White	Black	Asian
Walsh (2016)[244]	CAP	98%	<2%	<2%
Thompson (2006)[245]	PCPT	95.6%	3.2%	NR
Pinksy (2016)[62]	PLCO	85.0%	4.4%	4.0%
Eldred-Evans (2020)[246]	IP1-PROSTAGRAM	39%	33%	23%

Abbreviations: Not reported (NR)

Excludes ERSPC and STHLM3 which did not report ethnicity data.

A similar tabularised comparison between studies could not be completed for socioeconomic deprivation due to geographic and temporal differences in socioeconomic classification. Both PLCO[234] and CAP[235] have reported that participants were skewed towards men of a higher socioeconomic status. In IP1-PROSTAGRAM cohort, there were more men from a lower socioeconomic group due to the influence of the targeted recruitment strategy.

It is interesting that the direct mail strategy was shown to be effective at recruiting an ethnic and socioeconomic distribution commensurate to the demographics of the local population. The primary care practices were based in West London which was selected for its wide and well documented ethnic diversity and variation in socioeconomic deprivation. Previous studies have also found that a well-designed and carefully researched postal recruitment strategy can accrue a reasonable uptake of harder to reach populations[247].

The higher levels of socioeconomic deprivation in the local area could be a factor in the lower response rates to postal invitations seen in this study (22.7%). Table 16 compares the response rates in IP1-PROSTAGRAM to previous screening studies showing the difference in response rate between this study and previous population based screening trials[244, 248-250]. The reasons for non-participation have not been specifically explored in this study. It is possible that factors such as the repeated invitations, considerable infrastructure support, and delivery of a single non-invasive screening test could have contributed to the response rates in these large national trials.

Table 16: Response Rates to cancer screening invitations

Author (Year)	Study	Response	Invites	Response Rate
Roobol (2013)[248]	ERSPC Rotterdam	42,376	88,283	48.0%
Walsh (2016)[244]	CAP	90,300	197,763	45.6%
Gronberg (2015)[249]	STHLM3 Training	11,130	32,823	33.9%
Gronberg(2015)[249]	STHLM3 Validation	47,688	111,819	42.6%
Field (2016)[250]	UKLS (lung cancer)	75,958	247,354	30.7%
Eldred-Evans (2020)[246]	IP1-PROSTAGRAM	311	1370	22.7%

In IP1-PROSTAGRAM, due to the rapid recruitment from other recruitment strategies, the study completed before Primary Care Practices could send second invitations to non-responders. Previous population-based screening studies have increased response rates from sending reminder invites with scheduled appointments[240]. In cervical cancer screening a reminder letter with a pre-booked appointment increased participation two-fold compared to a single open invitation[251].

In addition, as fast MRI is a new and previously untested screening modality, there was an ethical imperative to emphasise this within the recruitment materials. In contrast, many large population screening randomised trials evaluating established screening tests have used a single-consent Zelen design. In this design, participants are randomised to a trial arm before

consenting to participate and informed consent is only obtained after randomisation while members of the control arm do not give consent.

The final reason for the lower response rates could be related to the fact that participants for IP1-PROSTAGRAM were being recruited to a trial offering multiple screening modalities and in particular a more invasive transrectal ultrasound. It is well-established that there are many psychosocial barriers to prostate cancer screening particularly in certain ethnic groups and especially where it involves rectal procedures[252]. For this reason, the acceptability of fast MRI as a screening test forms a critical component of the evaluation in this thesis and will be explored in more detail in Chapter 8.

5.5.3 Implications of Findings

This experience of delivering a targeted recruitment strategy within IP1-PROSTAGRAM underscores that the success of recruiting under-represented groups requires careful planning to design strategies which promote the trial to the relevant target population. The recruitment strategy for IP1-PROSTAGRAM was developed in the year prior to trial launch and the proportions of black men and lower socioeconomic groups recruited exceeded pre-trial expectations.

Future studies should consider setting a priori minority accrual goals and considering which recruitment strategies, either individually or in combination, can deliver those targets. In our experience, where a carefully designed direct mail strategy was delivered in an area of high ethnic diversity, it appeared to be successful at recruiting a group of men which was broadly representative of the local population. Direct mail had the additional advantage of providing a steady and predicable flow of responses which was useful for planning downstream trial infrastructure. In contrast the targeted and media recruitment strategies had large and less predictable peaks in response rates which can overwhelm trial infrastructure and cause delays in responding to potential participants.

The finding that direct mail may be the optimal approach highlights the importance in selection of recruitment sites for screening trials accounting for the potential for minority recruitment in each area. It is acknowledged that it may not be practical for a large population-based screening study to select recruitment areas with only diverse populations. For example STHLM3, was delivered in Sweden, a country with low rates of ethnic diversity compared to the rest of the world. Due to geographical constraints it can be predicted that clinical trials will lead to under-recruitment of certain demographics. In such cases the use of

a targeted strategy, similar to the approach described here and in other publications, would be a useful supplementary recruitment method to enhance recruitment and ensure a balanced study population.

If targeted recruitment is used in future studies with a different design and outcome to IP1-PROSTAGRAM, it is recommended that accrual goals are set a priori during protocol development. These targets should be based on racial and ethnic proportions in the UK, USA and Europe as regions where prostate cancer screening is most likely to be implemented. The study protocol also needs to be designed to allow on-going monitoring of the demographics of participants to avoid unplanned over-recruitment.

A similar process was performed in IP1-PROSTAGRAM and led to us identifying during the interim review that there was a need to balance enrolment between different recruitment methods to minimise over-representation of black men within the study population. This was necessary due to the high number of responses from targeted recruitment (n=387). In response recruitment from this group was restricted to 131 men (32%) meaning the majority of men were recruited from the direct mail strategy (n=219).

5.5.4 Limitations

A number of limitations should be acknowledged. First, data governance and ethical regulations precluded the sharing of the ethnicity data of non-participants. Consequently, the finding that the direct mail achieved a representative recruitment rate of black men is based on a comparison of direct mail participants (8% black men) and local population (9% black men). In addition, the recruitment materials were not translated into different languages and a number of potential participants were excluded due to insufficient English for consent. This was a consequence of funding constraints as had we recruited participants with insufficient English to consent and understand trial processes, translators would have been needed at multiple points in the trial.

Second, it is acknowledged that at the individual primary care practice level there are wide differences in demographics and, without access to the ethnicity of non-responders, it is uncertain whether ethnicities of direct mail participants are a consequence of the sample of men who received letter. Equally, while responders and non-responders were similar in terms of deprivation level, it cannot be excluded that other important factors such as education, marital status or household income could have influenced this as has been shown in other publications[253].

Third, the response rates for the majority of recruitment methods could not be calculated due to the nature of strategies such as word-of-mouth or posters. Therefore it was unknown how many men heard about the study via these recruitment methods. In addition, the response rate to letters is known to be underestimated as it cannot confirm the number of the target population who actually received and read the invitation letters. Inaccuracies or incomplete addresses remain an issue in medical records and it is likely that a number of letters were undelivered and did not reach the intended recipient. As an alternative metric, the cost per participant expressing an interest was calculated which provides insight into the cost-utility of each screening method.

Fourth, the primary care practices self-selected whether to send invitations via letter or SMS. This non-randomised design has inherent bias and does not account for differences in the population of the family practices. Despite the non-randomised design, this chapter has reported similar results to previous studies which found a lower uptake via text message recruitment compared to invitations by letter[254].

Fifth, the costs associated with each recruitment method included only direct costs and excluded personnel and overhead costs. Cost data for certain methods, such as social media recruitment, could not be estimated as there were no direct costs and no data estimates for staff time were available.

5.6 Conclusion

The participation of minorities is essential to ensuring results of screening trials are generalisable across the population. The findings of this chapter suggest that, where invitations materials have been designed to engage a diverse population, it is possible to achieve a representative uptake (including black men and those from a lower socioeconomic group) from direct mail recruitment. Our data supports the use of direct mail as an effective method of recruiting to population-based screening trials. The targeted recruitment strategy was also a viable method of recruiting high-risk individuals and would be recommended for use in combination with direct mail in geographic areas where certain groups are unrepresented in the target population.

Chapter 6 – Population-based prostate cancer screening using a prospective, blinded, paired screen-positive comparison of PSA and fast MRI

6.1 Overview

This chapter will report the primary outcomes of the IP1-PROSTAGRAM trial. The study was designed to determine the appropriate threshold denotating a screen-positive fast MRI if used as an exclusive screening test. The primary outcome compared the positive test rate of different MRI thresholds and the standard PSA threshold. This chapter forms the basis of work that I have presented at ASCO 2020[vii] and published in JAMA Oncology[viii].

6.2 Introduction

Any new screening test requires setting an appropriate threshold to denote a screen-positive result and this has been highlighted by the UK National Screening Committee which states that screening tests must have "a suitable cut-off level defined and agreed"[1]. The fast MRI protocol described in Chapter 4 suggested that there was a trade-off between two potential thresholds to denote a screen-positive MRI.

In Chapter 4, a threshold MRI score ≥ 3 had a high sensitivity for significant disease but led to an excessive biopsy rate and detection of equivocal lesions. An MRI score ≥ 4 generated fewer biopsies and equivocal lesions but the trade-off was missing significant cancer. The study was conducted in a high prevalence setting and could not draw conclusions on which threshold would be appropriate for a low prevalence screening population.

The aim of IP1-PROSTAGRAM was to evaluate the performance of both these thresholds in men aged 50 to 69 years in the general population. Prior to this study there had been no prospective clinical trials evaluating the performance of a fast MRI protocol across

[vii] Eldred-Evans, D., et al. (2020). "Population-based prostate cancer screening using a prospective, blinded, paired screen-positive comparison of PSA and fast MRI: The IP1-PROSTAGRAM study." Journal of clinical oncology **38**(15_suppl): 5513-5513.

[viii] Eldred-Evans, D., et al. (2021). "Population-Based Prostate Cancer Screening With Magnetic Resonance or Ultrasonography: The IP1-PROSTAGRAM study." JAMA Oncology (in press)

thresholds. In this study, the performance of each threshold is compared across a variety of metrics including biopsy rate, cancer detection rates and interobserver variability.

6.3 Methods

6.3.1 Study design and participants

IP1-PROSTAGRAM was a prospective, population-based, blinded cohort study conducted in the UK. Men were invited to participate using recruitment methods which were described Chapter 5. Inclusion criteria were men aged 50 to 69 years with a life expectancy of at least 10 years. Exclusion criteria were a PSA or prostate MRI in the previous two years, a urinary infection or prostatitis in the previous six months, a history of prostate cancer or any contraindication to MRI or prostate biopsy. To minimise attrition bias, participants were blinded to the screening test results until study completion.

6.3.2 Procedures

6.3.2.1 MRI

All participants who met the eligibility criteria underwent a fast MRI scan which was developed following the sequences described in Chapter 5 with reference to consensus guidelines[201, 255]. The full acquisition parameters and time for each sequence are listed in Table 17 with an example of a screen-positive fast MRI in Figure 27. All examinations were performed on a 1.5T (Siemens Magenetom Aera) or 3T (Siemen Magenetom Verio syngo MR B17) system using a standard phased-array body coils. The protocol time of the 3T protocol was 14 mins 17 seconds and for the 1.5 protocol was 15 mins 42 seconds.

The MRI scans were performed at two sites and each site's scans were centrally reviewed for quality prior to the start of the study. The quality assurance (QA) process required all manufacturer's service summaries to be sent to the IP1-PROSTAGRAM team before the beginning and end of the study.

The MRI scans were independently assessed by two reporters blinded to the PSA as well as the demographic and clinical information apart from age. The primary reporters included two uro-radiologists with eight and nine years of experience and a specialist interest in prostate cancer imaging. There was a centralised meeting prior to starting the study where the readers agreed a standardized screen reading protocol.

Table 17: Fast MRI Sequence Details

Sequence	Plane	TR (ms)	TE (ms)	Aver-ages	FA (degree)	WFS (pix)	BW (Hz/Px)	FoV (mm)	Phase FOV (% of FOV)	Over-sampling (% of FOV)	Phase enc. direction	Slice thickness (mm)	Slice gap (% of slice thickness)	TSE/EPI factor	FS method	Matrix	Phase res. (% of matrix)	Recon. voxel size (mm)	Sequence duration (mm:ss)
3T SIEMENS MAGNETOM Verio syngo MR B17																			
Localiser	Multiplanar	1000	92	1	150	1.2	349	400	100	20	Multiple	7	100	256		256	100	1.6x1.6x7	00:15
T2 TSE	Sagittal	7000	101	3	Min 150	2.0	200	200	100	43	H>F	3	20	25		320	80	0.8x0.6x3	02:57
T2 TSE	Axial	7000	108	2	Min 150	1.1	200	363	100	100	R>L	3	0	24		320	80	0.8x0.6x3	02:43
DWI (b0, 150, 400, 1000)	Axial	8500	80	3		0.2	1698	250	100	30	A>P	3	0	128	SPAIR	128	100	2 x 2 x 3	04:42
DWI (b1500)	Axial	9100	85	7		0.2	1698	250	100	30	A>P	3	0	128	SPAIR	128	100	2 x 2 x 3	03:40
1.5T SIEMENS MAGNETOM Aera																			
Localiser	Multiplanar	1000	93	1	180	0.4	501	400	100	20	Multiple	7	100	256		256	100	1.6x1.6x7	00:11
T2 TSE	Sagittal	5280	125	3	Min 150	1	200	200	100	100	H>F	3	20	23		320	80	0.6x0.6x3	03:17
T2 TSE	Axial	4590	135	3	Min 150	1	200	200	100	100	R>L	3	0	23		320	80	0.6x0.6x3	02:51
DWI (b0, 150, 400, 1000)	Axial	7500	67	2, 3, 4, 5		0.1	1507	250	100	30	A>P	3	0	128	SPAIR	128	100	2 x 2 x 3	05:23
DWI (b1500)	Axial	7500	68	9		0.1	1502	250	100	30	A>P	3	0	128	SPAIR	128	100	2 x 2 x 3	04:00

All scans were performed with intravenous administration of 20 mg hyoscine butylbromide. If contraindicated, 1 mg of glucagon hydrochloride was used intravenously. If both bowel relaxants were contra-indicated, no medication was used

Figure 27: A screen-positive fast MRI

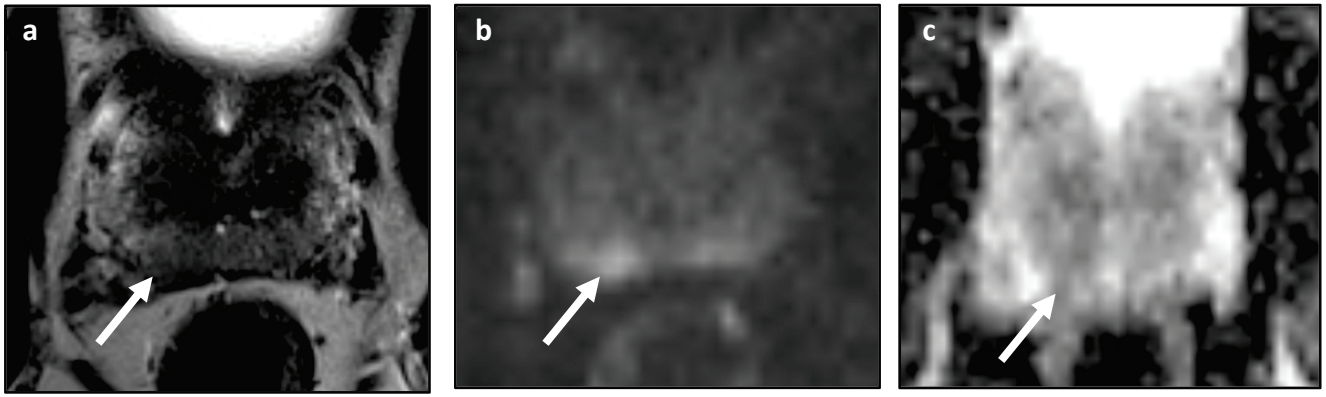


Figure: A selected case showing a screen-positive fast MRI protocol. This case shows a 57-year-old with PSA 1.02. He had no risk factors for prostate cancer and a benign DRE.

A biparametric MRI (bpMRI) showed a basal right peripheral zone lesion with restricted diffusion on DWI (b) and corresponding hypointense signal on ADC (c). The lesion was score 4 out of 5 on PI-RADS v2 and Likert scales. A targeted biopsy revealed Gleason 3+4 in all targeted cores with maximum cancer core length 7mm.

All scans were independently reported using a scale from 1 to 5, with higher numbers indicating greater likelihood of clinically significant prostate cancer. The MRI scoring used the Prostate Imaging–Reporting and Data System, version 2 (PI-RADS v2)[18] in accordance with the European Society of Urogenital Radiology (ESUR) PI-RADS Version 2.1 without DCE. This sets out the scoring guidelines when bpMRI is performed and DCE data is not obtained.

In this situation, the PI-RADS assessment category for a finding in the PZ is based solely on the DWI score and TZ assessments remain unchanged. Radiologists were also able to provide a Likert score for each lesion, which is a subjective assessment of the likelihood of significant disease. The score is based on likelihood of clinically significant cancer as defined on pathology/histology as Gleason score ≥ 7 (including 3+4 with prominent but not predominant Gleason 4 component) and/or volume ≥0.5cc and/or extra prostatic extension (EPE).

To assess interobserver agreement, 20% of the MRIs, stratified by MRI PI-RADS score to ensure a representative sample of each score, were selected at random from both sites and reviewed by an independent, blinded third radiologist with over 15 years' experience in prostate MRI. All men who were upgraded by the third independent reader were reviewed clinically provided they had not already undergone a biopsy as part of the trial. All scans were reviewed at the reporting radiologist's site on a picture archiving and communication system (PACS). Full details of each reader's level of experience is shown in Table 18.

Table 18: MRI Reporter Experience

	Radiologist 1	**Radiologist 2**	**Radiologist 3**
Role in study	Primary reader	Primary reader	Secondary reader
Number of years' experience	9	8	> 15 years
Number of prostate MRIs per annum	1,000	300	300

6.3.2.2 PSA screening

All participants also underwent a PSA test. This was obtained prior to the MRI scan at initial screening visit. The PSA samples were processed and tested at a single laboratory (Imperial College NHS Health Care Trust, London, UK). Samples were processed on the same day of collection. Serum PSA was measured using an automated chemiluminescent microparticle immunoassay analyser (Abbott Diagnostics, Abbott Park, IL, USA) referenced against the

World Health Organization (WHO) First International Standard for PSA (90:10), coded 96/670[21].

All participants were reviewed by a urologist to exclude those with a urinary tract infection, perform a digital rectal examination (DRE) and conduct a transrectal ultrasound (TRUS) examination as part of the secondary outcome of the study. The DRE and TRUS was only performed after the blood sample for PSA testing had been collected. The clinician was required to follow a standardised Urinary Tract Dipstick Algorithm when deciding whether to perform urine analysis and/or MSU. A summary of this algorithm is shown in Figure 28.

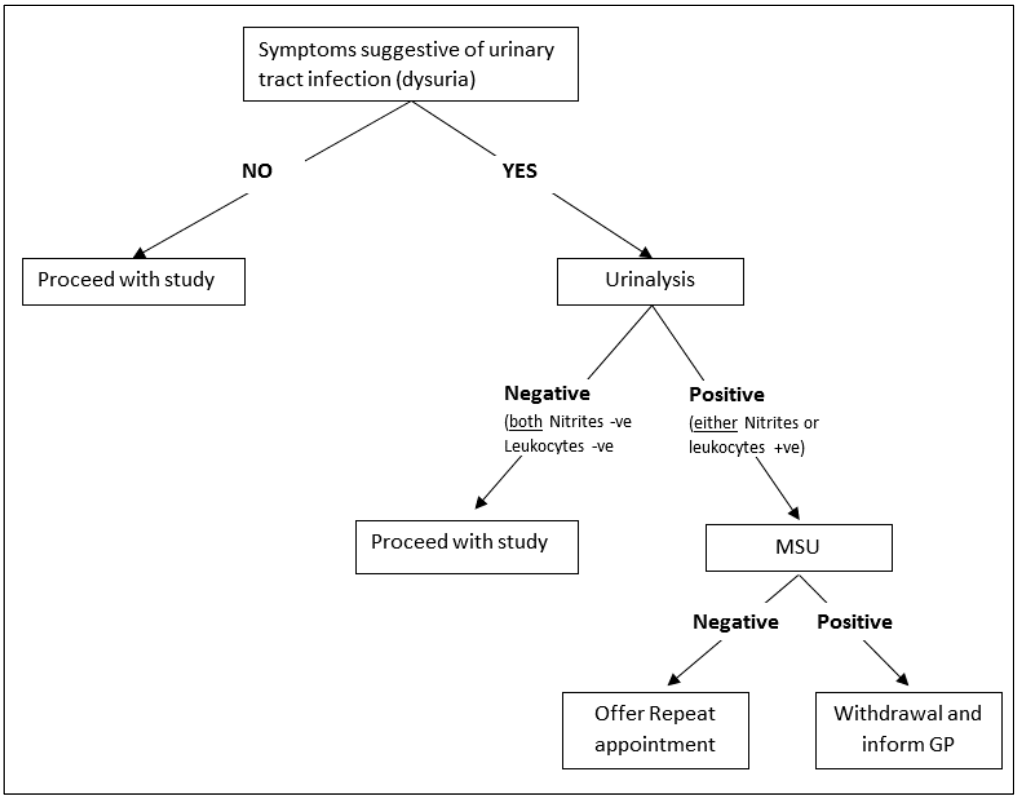


Figure 28: UTI Decision algorithm to reduce false positive PSA tests.

6.3.2.3 Biopsy

A biopsy was indicated in the presence of any screen-positive test results. In addition to fast MRI and PSA, the screening protocol included an additional new imaging test, shear wave elastrography (SWE). Participants were blinded to the indication for biopsy until the biopsy procedure was completed. The biopsy was performed within 12 weeks of the screening visit according to a standardised protocol. All biopsies were performed using a biplanar transrectal ultrasound (Hitachi Prerius, Hitachi Medical Corporation, Tokyo, Japan) mounted to a brachytherapy stepper and grid (Civco, Kalona, IA, USA). Image-fusion biopsies were carried out with software-assisted registration (BiopSee, MedCom GmbH, Darmstadt,

Germany). All biopsies were carried out at the lead site by operators who were independent of the index test result. Operators were urologists experienced in transperineal prostate biopsy and all operators underwent additional training in the study-specific biopsy protocol.

The procedure could be carried out under local anaesthetic +/- conscious sedation depending on patient-physician preference. In both approaches, bupivacaine 0.5% with adrenaline 20ml was infiltrated into skin and around the upper part of the anal verge followed by further infiltration of 2-3ml into deeper tissue around where biopsy needles are likely to penetrate. Following the superficial infiltration, a peri-prostatic block was completed using bupivacaine 0.5% and lidocaine 1% mixture used as peri-prostatic block under ultrasound guidance.

Additional image-fusion targeted biopsies of all MRI and ultrasound lesions were also carried out with individual targets potted and reported separately. If both MRI and ultrasound lesions were present, the order of targeted biopsy was allocated by computer-generated random number sequence and block randomization. A maximum of six imaging targets could be included in the IP1-PROSTAGRAM biopsy standard operating procedure. The targeted procedure required a minimum four to six targeted cores for each target scored 3, 4, or 5. It was accepted that smaller areas could require a higher biopsy density to reduce risk of sampling error. Following image targeting, contralateral PZ sampling of imaging-negative areas was carried out. In the event no imaging targets were available then non-targeted systematic biopsy of bilateral peripheral zones were taken.

Reporting radiologists were not present during the biopsy to ensure the reference test remained independent of the index test. All biopsies were centrally reviewed by expert uropathologists according to the International Society of Urological Pathology (ISUP) guidelines[20].

6.3.3 Outcomes

The primary outcome of IP1-PROSTAGRAM was a head-to-head comparison of the proportion of screen-positive PSA and fast MRI at the two pre-specified thresholds. Other secondary outcome measures included the percentage of false-positive results and the proportion of men with clinically significant and clinically insignificant cancer detected by each screening test. These outcomes will allow an evaluation of the appropriate MRI threshold in the general male population aged 50 to 69 years.

A screen-positive fast MRI was pre-defined and each MRI was scored on a whole gland level and dichotomised at a score of 3 (equivocal) and 4 (likely clinically-significant cancer) to create two thresholds, a score equal to or above which defined a screen-positive fast MRI. This allowed evaluation of an optimal threshold to use in a future definitive screening study. The PSA level was dichotomised as screen-positive (≥3ng/ml) or screen-negative (<3ng/ml).

Clinically significant cancer was defined across multiple pre-specified ordinal histological disease classes. The primary definition was Gleason score >/=3+4 or greater (ISUP >/=2) as this has become a more universally agreed histological definition of clinically significant cancer. Data are also presented on other definitions such as Gleason ≥ 3 + 4 and/or maximum cancer core length ≥ 4mm (Definition 2), Gleason ≥ 3 + 4 and/or maximum cancer core length ≥ 6mm (Definition 3) and Gleason ≥4 + 3 or greater (ISUP>/=3).

The frequency and incidence of adverse events (AEs) and serious adverse events (SAEs) occurring through the course of the study was assessed. An adverse event (AE) is any untoward medical occurrence in a patient or clinical trial subject. AEs were recorded as any unfavourable and unintended sign or symptom, whether or not they were considered to be related to screen protocol. Serious adverse events (SAEs) were recorded throughout the study.

6.3.4 Statistical Analysis

6.3.4.1 Sample size

The study was powered for the primary objective to determine the prevalence of screen-positive MRIs in the general male population aged 50-69 years. The sample size calculations followed the formula recommended by Naing et al[256] to determine an adequate sample size to estimate the prevalence of screen-positive MRIs with a precision of +/- 5%:

$$n = \frac{Z^2 P(1-P)}{d^2}$$

Z = Z statistics for a level of confidence, P = expected prevalence or proportion and d = precision.

The sample size calculation requires an estimate of the prevalence of screen-positive MRI (p). There are no previous studies that provide a reliable estimate of this figure in men aged 50-69 years at average risk of prostate cancer. We have estimated this figure based on a number of assumptions, which are listed below. We have split the population into two groups depending on PSA level, as there are different levels of evidence for each group.

Group 1 (PSA≥3ng/ml): We have assumed this to be 73% based on a combination of studies which included the PROMIS study which has shown that in a group of biopsy-naïve men referred with a suspicion of prostate cancer, the prevalence of positive MRI (Likert ≥ 3) was 72.6%[88]. The PRECISION study which used PI-RADS v2 found a prevalence of 71.1%[133] for PI-RADS ≥ 3. A systematic review which did not include PROMIS and PRECISION, and which categorised PI-RADS thresholds across different groups of men confirmed that 73% of biopsy-naïve men have a positive scan defined as PI-RADS score [257].

Group 2 (PSA<3ng/ml): There is limited data on the number of positive MRIs in this group so we have combined previous research estimating;

1. The prevalence of expected significant cancers:

   In the Prostate Cancer Prevention Trial (PCPT) this estimated that the prevalence of clinically significant disease in a normal PSA population is 2.2%[86, 258]. The reference test was a 6-core (sextant) biopsy which is known to underestimate the presence of cancer and has been replaced with a 10-12 core approach. However we do not have

any reference to estimate by how much the 6-core biopsy underestimates the proportion of clinically significant disease in a population with PSA <3.0. Therefore, we hypothesis that the prevalence of undiagnosed clinically significant disease in this group is 2.2%.

2. The performance characteristics of MRI:

   We will assume that the performance characteristics (sensitivity and specificity) of mp-MRI to detect this clinically significant disease are the same in a normal and raised PSA population. These performance characteristics are variable across in the literature. A meta-analysis by Rooij et al 2014[259] reported a sensitivity and specificity of 74% and 88% respectively and the recent PROMIS study[88] (not included in Rooij et al) reported a sensitivity and specificity of 93% and 41%. We have calculated the assumed prevalence of a positive MRI in a normal PSA population using both these performance characteristics using the following 2x2 tables

Table 19: Sensitivity & Specificity as per Rooij et al 2014[259]

		Clinically significant PCa		Total
		Diseased	Non-Diseased	
MRI	Positive	1.63%	11.74%	**13.37%**
	Negative	0.57%	86.06%	86.63%
	Total	2.20%	97.80%	

Therefore given these performance characteristics the prevalence of a positive MRI in a normal PSA population will be either 13.37% based on Rooij et al 2014[259] or 59.75% based on Ahmed et al 2017[88]

Table 20: Sensitivity & Specificity as per Ahmed et al 2017[88]

		Clinically significant PCa		Total
		Diseased	Non-Diseased	
MRI	Positive	2.05%	57.70%	59.75%
	Negative	0.15%	40.10%	40.25%
	Total	2.20%	97.80%	

The final part of the calculation is to combine the assumed prevalence of positive MRI in the normal and raised PSA groups to estimate the prevalence of a positive MRI in a mixed population. There is high quality evidence for the expected percentage of normal and raised PSA from the Cluster randomised trial of PSA testing for Prostate cancer (CAP)[260] that shows a raised PSA will occur in 10.4% and a PSA < 3ng/ml in 89.6%.

Therefore, using our estimates above for the prevalence of positive MRIs in a normal PSA population, we expect 12.0% (13.4% of 89.6%) from Rooij et al 2014[259] and 53.54% (59.8% of 89.6% from Ahmed et al 2017[88]. The positive prevalence in a raised PSA population is 7.6% (73% of 10.4%). This produces an assumed prevalence of positive MRI in both groups of 19.6% (Rooij et al[259]) or 61.1% (Ahmed et al[88]).

Using the formula by Naing et al[256], assuming a prevalence of 19.6% requires a sample size of 243 participants, while assuming a prevalence of 61.1% will require a sample size of 366 participants. Allowing for a 10% dropout a minimum target of 270 participants (low prevalence of screen-positive MRI) and a maximum target of 406 participants (high prevalence of screen-positive MRI) was calculated to provide a 5% precision estimate at a two-sided significance level of 0.05.

6.3.4.2 Primary outcome

The primary analysis included all participants who were enrolled and completed at least one screening test. The proportion of men with a screen-positive fast MRI and PSA was estimated with the corresponding 95% confidence interval (CI). Confidence intervals were calculated on logit-transformed estimates and then transformed back. Comparisons of proportions of screen-positive and screen-negative results, between pairs of PSA and MRI was conducted using McNemar chi square tests. The McNemar test for paired proportions assessed whether there was any marginal homogeneity of results between pairs of screening tests at each threshold. The analysis was repeated for screen-positive PI-RADS and Likert scores at the pre-defined thresholds. All tests were two tailed with a 5% significance level.

6.3.4.3 Secondary outcomes

The proportions of false positive results were calculated for each screening test with 95% CIs. The proportions of false positive was defined as a screen-positive result when significant prostate cancer was not present on biopsy. 2x2 contingency tables were reported to show the head-to-head differences in detection of clinically significant cancers stratified by method

of detection (targeted vs. non-targeted biopsy). The detection rates for each screening test were compared using McNemar chi square tests for paired proportions.

Continuous demographic variables were summarized for all participants included in the primary outcome with descriptive statistics (number, mean, median, SD, range, and 95% two-sided confidence limits). A STARD compliant flow charts will show the outcomes of all participants in the trial and adverse events summarized in a frequency table. The proportion of withdrawals for each screen-positive test were compared with an exact test for two-sample binomial proportions. The proportion of significant cancers as well as insignificant cancers over-diagnosed per 100 men screened, alongside positive predictive value were compared using the same method.

Interobserver agreement was assessed on a random sample of 20% of the MRI scans. The random sample was stratified by PI-RADS score to ensure an adequate mix of screen-positive and screen-negative results. The groups were negative (a score of 1 or 2), intermediate (a score of 3), or positive (a score of 4 or 5). A 5x5 contingency table was used to summarize the agreement in PI-RADS scores between the primary reporters and third independent reader.

Interobserver agreement was calculated by score agreement and concordance with biopsy decision at the pre-defined screen-positive thresholds using PI-RADS score ≥ 3 or ≥ 4. To address the challenge of calculating interobserver agreement in a low prevalence setting, interobserver agreement was measured using four methods: percentage agreement, Cohen's Kappa index, interclass correlation coefficient using the average fixed raters (ICC3k) and Gwet's first-order agreement coefficient statistic (AC1)[261] with 95% confidence intervals.

Although Kappa is one of the most common agreement statistics it assumes that agreement is at random, so it expresses agreement beyond the random expected level agreement. In contrast Gwet's AC1 statistic assumes that agreement between scores are not completely random. It accounts for the fact that in some cases there will be an easy agreement and in other cases the agreement will be difficult. These were interpreted using the scale proposed by Altman with the strength of agreement as poor, fair, moderate, good or very good for values exceeding 0, 0.2, 0.4, 0.6, and 0.8, respectively[262].

6.4 Results

6.4.1 Study Population

A total of 411 men attended for screening and 408 men were eligible to receive a fast MRI and PSA test. The characteristics of these participants are shown in Table 21. The majority of participants (77.2%) had PSA and MRI on the same day (median 0 days, range 0-76). A total of 185 (45%) participants had a 'positive' finding from one or more of the screening tests and were advised to have a biopsy.

The rate of attrition (defined as withdrawing from the study following a screen-positive result and recommendation for biopsy) was 9.7% (18 of 185 [95%CI 10.7-28.4]). The majority requested withdrawal citing a wish to be unblinded to the reason for biopsy. There was no difference in the rates of withdrawal between men with screen-positive MRI or PSA results (9 of 98, 9.2% [95%CI 4.2-17.4] vs. 5 of 41, 12.2% [95%CI 4.0-28.4]; p=0.59). In total, 167 men had biopsy results available for the paired screen-positive analysis.

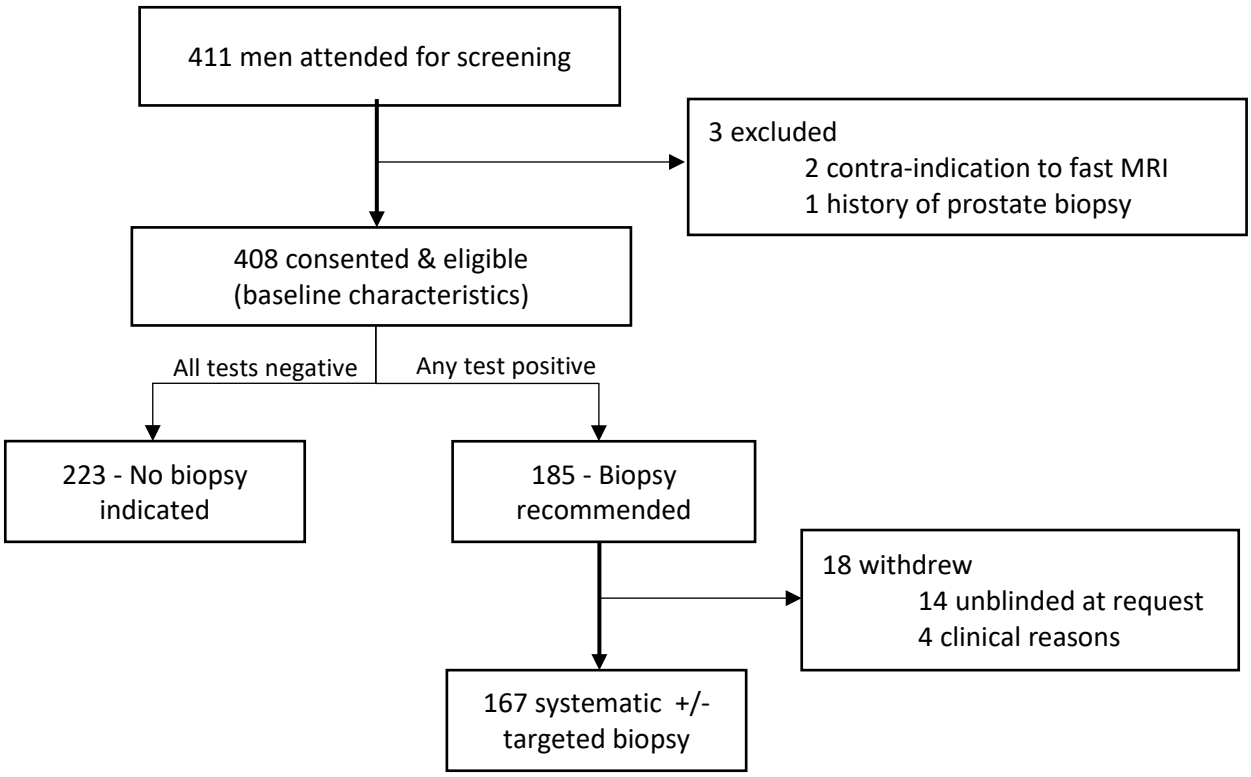


Figure 29: Flow of participants

Table 21: Baseline characteristics	
	Participants (n = 408)
Age (years)	
50-54	140 (34.3%)
55-59	127 (31.1%)
60-64	85 (20.8%)
65-69	56 (13.7%)
Ethnicity	
White	155 (38.0%)
Black	132 (32.4%)
Asian	94 (23.0%)
Other	18 (4.4%)
Mixed race	9 (2.2%)
Charlson Comorbidity	
0 (none)	324 (79.4%)
1 (mild)	49 (12.0%)
≥ 2 (severe)	13 (3.2%)
Unknown	22 (5.4%)
First degree relative with prostate cancer	
Yes	43 (10.5%)
No	360 (88.2%)
Unknown	5 (1.2%)
5 alpha reductase inhibitors	
Yes	1 (0.2%)
No	407 (99.8%)
IPSS score	
Mild (≤7)	277 (67.9%)
Moderate	112 (27.5%)
Severe	10 (2.5%)
Unknown	9 (2.2%)
Previous PSA test	
No	291 (71.3%)
Yes	
2-3 years ago	30 (7.4%)
3-5 years ago	41 (10.0%)
More than 5 years ago	26 (6.4%)
Date not known*	6 (1.5%)
Unknown	14 (3.4%)

Data are n (%).
*Exclusion criteria included any PSA in last 2yrs

6.4.2 Primary Outcome

The proportion with a screen-positive MRI defined as PI-RADS score ≥3 was higher than the proportion with a screen-positive PSA (72 of 406, 17.7% [95%CI 14.3-21.8] vs. 40 of 406, 9.9% [95%CI 7.3-13.2]; p<0.001). For PI-RADS score ≥4, this was similar to PSA (43 of 406, 10.6%, 95%CI 7.9-14.0 vs. 40 of 406, 9.9%, 95%CI 7.3-13.2; p=0.71). The proportion of screen-positive MRI (score >/=3) was also higher than the proportion with MRI (score >/=4) (72 of 406, 17.7% [95%CI 14.3-21.8] vs. 43 of 406, 10.6%, 95%CI 7.9-14.0; p<0.001).

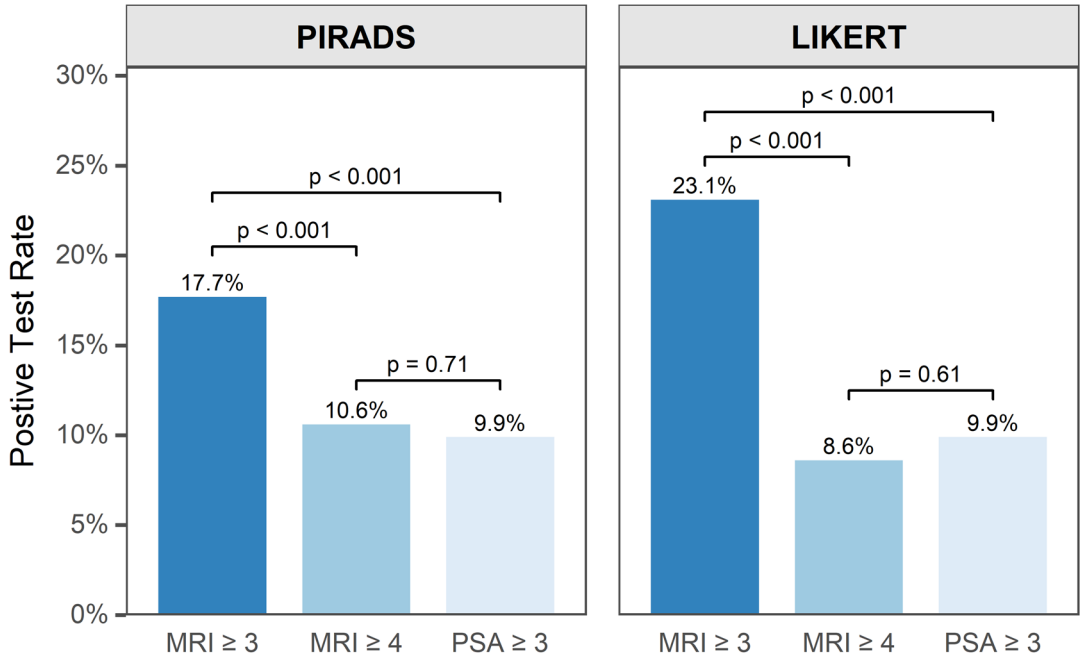


Figure 30: The prevalence of positive test results for PI-RADS, Likert and PSA at each threshold. The comparison of proportions between paired screen-positive test was conducted using McNemar chi square tests.

A comparison using Likert shows the differences were similar with Likert ≥3 being higher than PSA (94 of 406, 23.0% [95%CI 19.2-27.4] vs. 40 of 406, 9.9% [95%CI 7.3-13.2]; p<0.001). LIKERT ≥4 was similar to PSA (35 of 406, 8.6%, 95%CI 6.2-11.7 vs. 40 of 406, 9.9%, 95%CI 7.3-13.2; p=0.603). The proportion of Likert ≥3 scores was also higher than Likert ≥4 (94 of 406, 23.0% [95%CI 19.2-27.4] vs. 35 of 406, 8.6%, 95%CI 6.2-11.7 vs. 40 of 406, 9.9%, 95%CI 7.3-13.2; p=0.71; p<0.001).

The distribution of screen-positive PI-RADS scores was 7.1% (29 of 406 [95%CI 5.0-10.1]) with PI-RADS score 3, 7.8% (32 of 406 [(95%CI 5.6-10.9]) with PI-RADS score 4 and 2.7% (11 of 406 [95% CI 1.5-4.8]) with PI-RADS score 5. For screen-positive Likert scores there were 14.5% (59 of 406 [95%CI 11.4-18.2]) with Likert 3, 7.6% with Likert 4 (31 of 406 [95%CI 5.4-10.6]) and 1.0% (4 of 406 [95% CI 0.04-2.6]) with Likert 5.

A head to head comparison between the screen-positive score distribution showed a higher proportion of screen-negative PI-RADS scores compared to screen-negative Likert scores (334 of 406, 81.9% [95%CI 77.8-85.3] vs. 312 of 406, 76.5% [95%CI 72.1-80.3]; p<0.001) (Figure 31). There was a corresponding reduction in PI-RADS 3 scores compared to Likert 3 scores (29 of 406, 7.1% [95%CI 5.0-10.1] vs. 59 of 406, 14.5% [95%CI 11.4-18.2]; p<0.001). There was no significant difference in the proportion of PI-RADS and Likert score 4 (32 of 406, 7.8% [95%CI 5.6-10.9] vs. 31 of 406, 7.6% [95%CI 5.4-10.6]; p=1.00) or PI-RADS and Likert 5 (11 of 406, 2.7% [95%CI 1.5-4.8] vs. 4 of 406, 1.0% [95%CI 0.04-2.6]; p=0.69).

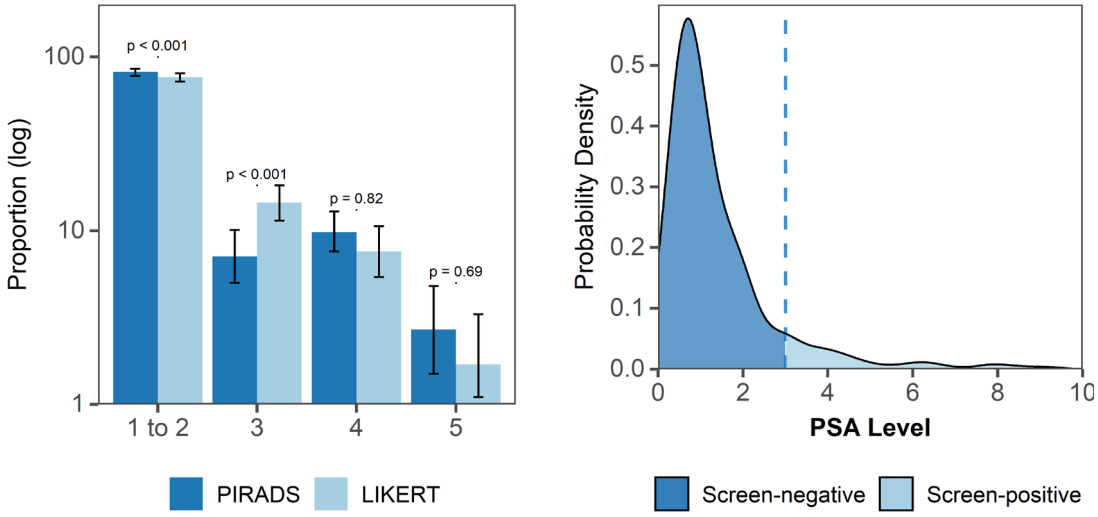


Figure 31: (A) The proportion of PI-RADS relative to Likert scores across each category. Bar charts includes 95% Cis and p values from McNemar Tests calculating matched pairs at each ordinal Likert or PI-RADS score (B): PSA distribution of participants. Constructed at 0.1ng/ml increments utilizing the corresponding PSA distribution histogram. The vertical access for the density plot represents an estimate of the distribution of PSA where density multiplied by 100 equals the percentage of participants with the PSA level. Area under the curves contains 100% of participants with a paired PSA and MRI result (n= 406). The results are dichotomised at 3ng/ml to represent the threshold used in IP1-PROSTAGRAM to denote a screen-positive vs. screen-negative PSA result.

The PSA frequency density plot demonstrates the distribution of PSA levels across all participants. The distribution of screen-negative PSA values was 0-1ng/ml for 217 of 406 (53.5% [95%CI 48.6-58.3]), 1-2ng/ml for 114 of 406 (28.1% [95%CI 23.9-32.6]) and 2-3ng/ml for 35 of 406 (8.6% [95%CI 6.3-11.8). For screen-positive values the distribution was 3-4ng/ml for 17 of 406 (4.2% [95%CI 2.6-6.6]), 4-5ng/ml for 9 of 406 (2.2% [95%CI 0.4-4.3), 5-10ng/ml for 12 of 406 (3.0% [95%CI 1.7-5.1) and ≥10ng/ml for 2 of 406 (0.5% [95%CI 0.01-1.7])).

6.4.3 Secondary Outcomes

6.4.3.1 PSA and MRI scores stratified by disease status

The distribution of PSA results and MRI Scores (PI-RADS) stratified by the presence and absence of significant cancer is shown in Figure 32. In those diagnosed with significant

prostate cancer, 82.4% (14 of 17) had a screen-positive MRI (Score 3-5) and 41.2% (7 of 17) had a screen-positive PSA (difference 41.2%, 95% CI 8.5-63.9, p=0.0158).

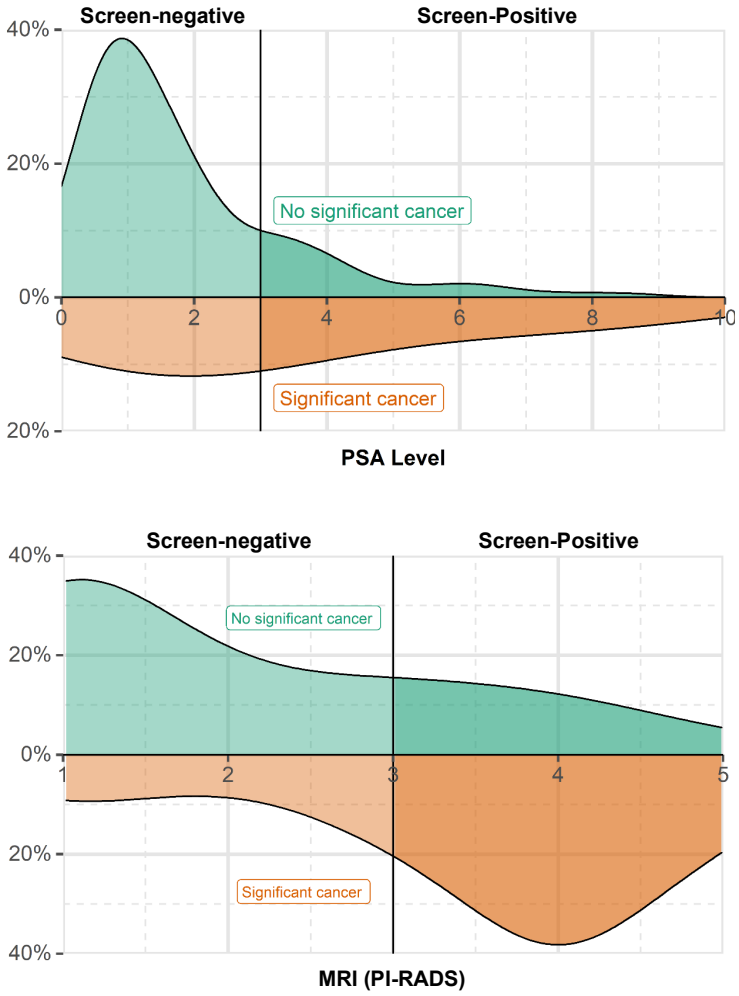


Figure 32: Density plot comparing the PSA and PI-RADS scores for men by disease status .Men with clinically significant prostate cancer are denoted in orange and men without significant cancer are shown in green. Clinically significant cancer defined by the primary definition (Gleason ≥ 3+4).

A screen-negative MRI (PI-RADS 1 or 2) was found in three men (17.6%) with significant disease and a screen-negative MRI (Likert 1 or 2) in two men (11.7%). At the higher threshold to denote a screen-negative MRI, there were six men (41.1%) with clinically significant disease who had a PI-RADS or Likert score 1, 2 or 3. The distribution of PSA scores in men with significant cancer included a PSA <=1ng/ml in three men, 1-2ng/ml in four men, 2-3ng/ml in three men and ≥3 in seven men. The median PSA in this group was 2.13ng/ml.

6.4.3.2 Interobserver Agreement

In total, 78 MRI scans underwent double-reporting by a third experienced radiologist who had no knowledge of PSA level, original MRI report or biopsy outcome. The individual scores are presented in Table 22. There were three participants with a screen-positive MRI (score 4-5) according to the independent third readers and these three participants were recalled for biopsy, all of which were benign

Table 22: Individual score agreement

		Secondary reader				
		1 – 2	**3**	**4**	**5**	**Total**
Primary reader	**1 - 2**	**21**	2	2	1	26
	3	14	**9**	2	1	26
	4	9	5	**5**	0	19
	5	2	1	1	**3**	7
	Total	46	17	10	5	78

The measurements of interobserver agreement is shown in Table 23. The percentage agreement for a biopsy threshold >/=3 threshold was 61.5% (95% CI 50.4- 71.5) and for >/=4 was 70.5% (95% CI 59.6-79.5). The percentage agreement increased with each PI-RADS score suggesting a higher level of agreement as the suspicion of cancer increases. The kappa statistic was fair at both threshold >/=3 and >/4. It increased to moderate agreement levels at higher thresholds (>=4) and PI-RADS score 5. The ICC and Gwett AC1 showed a similar trend although as these tests are less affected by prevalence the level of agreement was higher for thresholds score 4-5.

Table 23: Interobserver agreement by MRI threshold and Score

	Overall Agreement	Kappa (95% CI)	Gwett AC1 (95% CI)	ICC (95% CI)
Screen-positive Threshold				
MRI Score 3-5	61.5% (50.4- 71.5)	0.274 (0.07-0.48)	0.235 (0.01-0.46)	0.477 (0.18-0.67)
MRI Score 4-5	70.5% (59.6-79.5)	0.258 (0.03- 0.51)	0.518 (0.31-0.72)	0.427 (0.10-0.64)
PI-RADS Score				
3	67.9% (56.9-77.3)	0.211 (0.04-0.45)	0.466 (0.25-0.68)	0.358 (0.07- 0.59)
4	75.6% (65.1- 83.8)	0.213 (0.095-0.52)	0.651 (0.48-0.81)	0.363 (0.01-0.59)
5	92.3% (84.2-96.4)	0.462 (0.04-0.8)	0.910 (0.84-0.99)	0.632 (0.42-0.77)

Abbreviation: Interclass correlation coeffective using the average fixed rates score (ICC)

From the 78 scans selected for an independent report, 47 had a primary report from Reporter 1 and 31 from Reporter 2. A comparison between each primary reader and the third independent reader showed that for MRI >/3, percentage agreement between the two assessors varied from 59.6% (95% CI 49.3-69.8) to 64.5% (95% CI 55.5-75.3), kappa varied from 0.275 (95% CI 0.02-0.526) to 0.297 (95% CI 0.005-0.63), Gwett AC1 from 0.204 (95% CI 0.02-0.49) to 0.29 (95% CI 0.02-0.64) and ICC from 0.528 (95% CI 0.15-0.74) to 0.465 (95% CI 0.01-0.74). For threshold >/=4, percentage agreement varied between 72.3% to 67.7%, kappa from 0.3 to 0.21, Gwett ACC 1 from 0.56 to 0.45 and ICC from 0.521 to 0.357.

Table 24: Interobserver agreement between Reader 1 and Reader 2

	Overall Agreement (95% CI)	Kappa (95% CI)	Gwett AC1 (95% CI)	ICC (95% CI)
MRI Score 3-5				
Reader 1	59.6% (49.3-69.8)	0.275 (0.02-0.526)	0.204 (0.02-0.49)	0.528 (0.15-0.74)
Reader 2	64.5% (55.5-75.3)	0.297 (0.005-0.63)	0.290 (0.02-0.64)	0.465 (0.01-0.74)
MRI Score 4-5				
Reader 1	72.3% (61.4-82.5)	0.303 (0.02-0.625)	0.561 (0.31-0.81)	0.521 (0.14-0.73)
Reader 2	67.7% (57.2-75.2)	0.217 (0.01-0.616)	0.451 (0.11-0.79)	0.357 (0.02-0.69)

ICC = Interclass correlation coeffective using the average fixed rates score (ICC3k)

6.4.3.3 False Positive Rate

At the primary definitions of clinically significant disease (Gleason ≥ 3+4), there was a higher rate of false positive screening results for MRI at Score >/=3 compared to MRI at score >/= 4 (51 of 406, 12.6% [95%CI 9.7.-16.1] vs. 27 of 406, 6.7% [95% 4.6-9.5]; p=0.0044). The proportion of false positives for PSA was 7.1% (29 of 406, 95%CI 5.0-10.1). A selected case of a false-positive PSA result with a true negative MRI is shown in Figure 33.

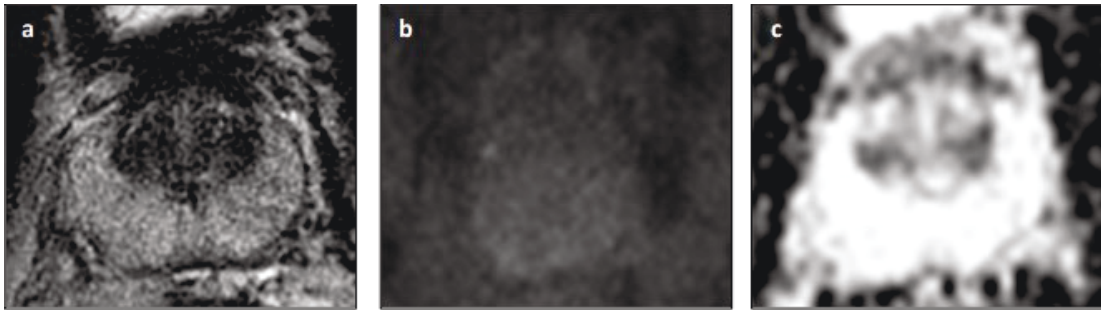


Figure 33: 55-year-old man with a screen-positive PSA 8.60ng/ml (a) Axial T2-weighted image (b) Diffusion-weighted imaging (b value = 1500 s/mm) (c) Apparent diffusion coefficient map through the mid-gland showing no abnormalities. The prostate volume was 34ml. The 12-core systematic biopsy did not identify any cancer.

For both secondary definitions of clinical significance on histology, the pattern was similar with a higher rate of false positive screening results MRI at Score >/=3 compared to MRI at score >/= 4 at both Definition 2 (50 of 406, 12.3% [95%CI 9.5.-15.9] vs. 26 of 406, 6.4% [95% 4.4-9.2]; p</=0.004) and Definition 3 (50 of 406, 12.3% [95%CI 9.5.-15.9] vs. 26 of 406, 6.4% [95% 4.4-9.2]; p=0.0039). The false positive rate for PSA at Definition 2 was 6.2% (25 of 406, [95%CI 4.2-8.9] and at Definition 3 was 6.7% (27 of 406, [95%CI 4.6-9.5] (p>0.05).

Table 25: False Positive Rate and Cancer Detection Rates

	MRI Score 3-5	MRI Score 4-5	PSA ≥ 3ng/ml
Primary Definition: Gleason ≥ 3 + 4 (ISUP ≥2)			
False Positive Rate	12.6% (9.7-16.1)	6.7% (4.6-9.5)	7.1% (5.0-10.1)
Cancer Detection Rate	3.4% (2.1-5.7)	2.7% (1.5-4.7)	1.7% (0.8-3.5)
Overdiagnosis Rate	1.7% (0.8-3.5)	1.2% (0.5-2.8)	1.5% (0.6-3.2)
Definition 2: Gleason ≥ 3 + 4 and/or maximum cancer core length ≥ 4mm			
False Positive Rate	12.3% (9.5-15.9)	6.4% (4.4-9.2)	6.2% (4.2-8.9)
Cancer Detection Rate	2.9% (1.7-5.0)	2.4% (1.3-4.4)	2.2% (1.2-4.1)
Overdiagnosis Rate	1.7% (0.8-3.4)	1.0% (0.3-2.5)	1.0% (0.3-2.5)
Definition 3: Gleason ≥ 3 + 4 and/or maximum cancer core length ≥ 6mm			
False Positive Rate	12.3% (9.5-15.9)	6.4% (4.4-9.2)	6.7% (4.6-9.5)
Cancer Detection Rate	2.2% (1.1-4.1)	2.2% (1.1-4.1)	1.5% (0.6-3.2)
Overdiagnosis Rate	2.9% (1.7-5.0)	1.7% (0.8-3.5)	1.5% (0.7-3.2)

Definitions: Any Gleason ≥ 4 + 3 (GrG ≥3) not included as 0 cases identified during study

6.4.3.4 Cancer detection rates

In total, 37 of 408 participants (9.1%) had prostate cancer on the reference test. There were 17 significant and 20 insignificant prostate cancers according to the Gleason ≥ 3+4 definition. For clinically-significant cancers, PSA (≥3ng/ml) detected seven cases, MRI (score 3-5) detected 14 cases and MRI (score 4-5) detected 11 cases. For clinically-insignificant cancers, PSA (≥3ng/ml) detected six cases, MRI (score 3-5) detected seven cases and MRI (score 4-5) detected five cases. Additional definitions of significant cancer are shown in Table 25.

The image-fusion targeted biopsy detected clinically significant cancer due to MRI (Score 3-5) in 14 cases and MRI (Score 4-5) in 11 cases. For systematic biopsy, clinically significant cancer was detected in four men where the PSA was ≥3ng/ml and there were no suspicious

areas on imaging. A breakdown of the numbers of significant and insignificant cancers detected by each biopsy modality is shown in Figure 34.

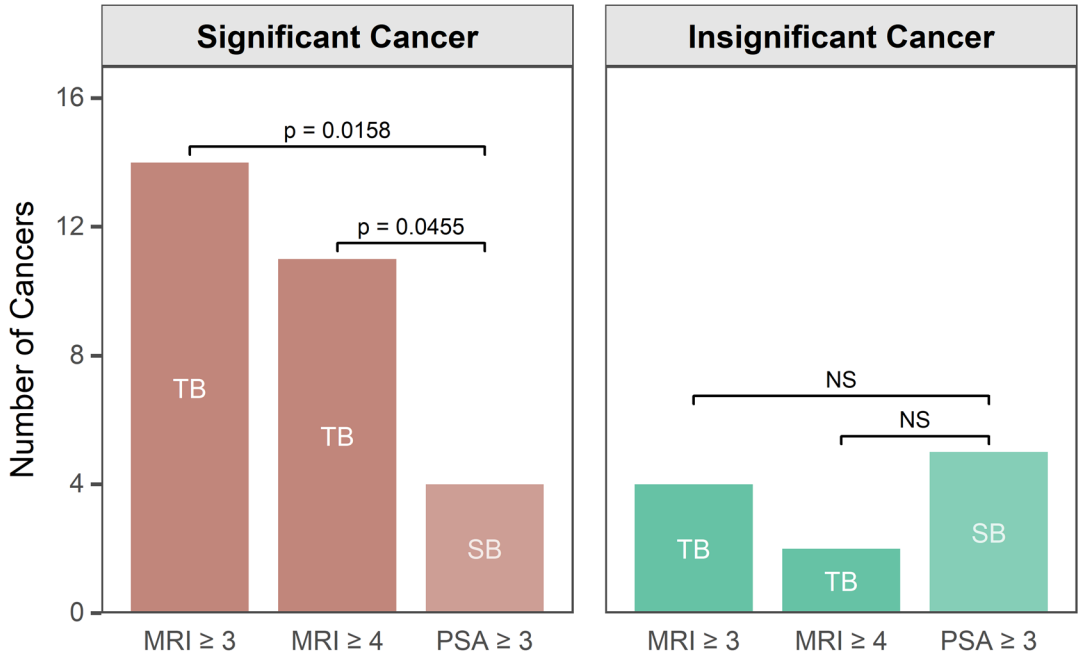


Figure 34: Number of significant and insignificant cancers detected by targeted biopsy (MRI) and systematic biopsy (PSA). Exact p values for McNemar's test for paired binary data.

The concordance and disconcordance of the significant cancers detected by each screening test is shown in Table 26. This presents two-way paired cancer detection data for the 17 clinically significant cancers detected. Using a targeted biopsy vs. systematic strategy, both MRI scores 3-5 and MRI 4-5 had significantly higher cancer detection rates than PSA ≥ 3ng/ml. There was no difference in detection of insignificant cancer (Gleason 3+3) between either test.

Table 26: Pairwise detected clinically-significant cancers by each screening test with reference standard either targeted (TB) biopsy, systematic biopsy (SB) or combined biopsy

A. Clinically significant prostate cancer detected at MRI score ≥ 3 vs. PSA ≥ 3 using targeted (TB) or systematic (SB) biopsy. p = 0.0158

	PSA ≥ 3	PSA ≥ 3	Total
MRI ≥ 3	3	10	13
MRI < 3	1	3	4
Total	4	13	17

B. Clinically significant prostate cancer detected at MRI score ≥ 4 vs. PSA ≥ 3 using targeted (TB) or systematic (SB) biopsy. p = 0.0455

	PSA ≥ 3	PSA < 3	Total
MRI ≥ 4	3	8	11
MRI < 4	1	5	6
Total	4	13	17

C. Clinically significant prostate cancer detected at MRI score ≥ 3 vs. PSA ≥ 3 using combined targeted and non-targeted biopsy. p = 0.0455					**D. Clinically significant prostate cancer detected at MRI score ≥ 4 vs. PSA ≥ 3 using combined targeted and non-targeted biopsy. p = 0.22**			
	PSA ≥ 3	PSA ≥ 3	Total			PSA ≥ 3	PSA < 3	Total
MRI ≥ 3	6	8	14		MRI ≥ 4	6	5	11
MRI < 3	1	2	3		MRI < 4	1	5	6
Total	7	10	17		Total	7	10	17
Exact p value for McNemar's test comparing paired binary data								

6.4.3.5 Treatment outcomes

Of the 37 prostate cancers which were identified, 20 (54%) were Gleason 3+3 with a maximum cancer core length (MCCL) of <2mm in 12 (60%) 2-4mm in 7 (35%), >4mm in 1 (5%). All Gleason 3+3 were treated with active surveillance which included 10 detected by MRI (Likert or PI-RADS ≥ 3), 5 detected by MRI (Likert or PI-RADS ≥ 4) and 6 detected by PSA.

Of the 17 Gleason 3+4 detected, the MCCL was <2mm in 3 (17.6%), 2-4mm in 3 (17.6%), 4-6mm in 2 (11.8%) and >6mm in 9 (52.9%). 8 patients commenced on active surveillance, 4 patients were treated with focal therapy, 2 patients had a radical prostatectomy and 3 patients had hormones and/or radiotherapy.

In the MRI (Score 3-5) group, 5 patients commenced on active surveillance, 4 patients treated with focal therapy, 2 patients had radical prostatectomy and 3 patients had hormones and/or radiotherapy. In the MRI (Score 4-5) group, 3 patients commenced active surveillance, 3 patients had focal therapy, 2 had a radical prostatectomy and 3 had hormones and/or radiotherapy. In the PSA group, 2 patients commenced on active surveillance, 1 patient was treated with focal therapy, 2 patients had a radical prostatectomy and 3 patients had hormones and/or radiotherapy.

Table 27: Treatment modality

	MRI Score 3-5 (N = 406)	**MRI Score 4-5** (N = 406)	**PSA ≥3ng/ml** (N = 406)
Active Surveillance	14 (3.4%)	6 (1.5%)	8 (2.0%)
Focal Therapy	4 (1.0%)	3 (0.7%)	1 (0.2%)
Radical Prostatectomy	2 (0.5%)	2 (0.5%)	2 (0.5%)
Hormones or Radiotherapy	3 (0.7%)	3 (0.7%)	3 (0.7%)

6.4.3.6 Adverse events

There were no serious adverse events during the study. The adverse events from the PSA and MRI screening tests were few and minor. The rate of at least one adverse event was 0.74% following MRI screening and 0.25% after PSA. Adverse events from MRI included procedure related anxiety in 2/408 (0.5%) and discomfort in 1/408 (0.25%). Adverse events from PSA included a superficial infection in 1/408 (0.25%).

The most frequent adverse events occurred secondary to the biopsy procedure and included haematuria in 5/408 (1.2%) and haematospermia in 4/408 (0.98%). A full list of adverse events is provided in Table 28.

Table 28: Adverse Events by procedure

	N = 408 (%)
MRI	
Procedure related anxiety/pain	2 (0.5%)
Sensation of over-heating	1 (0.25%)
PSA	
Superficial Infection	1 (0.25%)
Other study procedures (e.g. biopsy)	
Haematuria	5 (1.2%)
Haematospermia	4 (0.98%)
Procedure related pain	1 (0.25%)
Other	4 (0.98%)
Urinary Tract Infection	2 (0.5%)
Lower urinary tract symptoms	2 (0.5%)
Unrelated to study procedures	
Cold-like symptoms	1 (0.25%)

6.5 Discussion

6.5.1 Principle Findings

The primary aim of IP1-PROSTAGRAM was to assess the prevalence of screen-positive fast MRIs at different thresholds compared to PSA in order to determine an appropriate threshold to denote a screen-positive MRI. The secondary outcomes evaluated interobserver variability of fast MRI and compared false-positives, detection rate and overdiagnosis for each MRI threshold and PSA.

The results show that the positive test rate and false positive rate were significantly higher for MRI (Score 3-5) compared to MRI (Score 4-5) and PSA ≥ 3ng/ml. For MRI (Score 3-5),

17.7% were classified as screen-positive and 12.6% had a false positive result. This was due to the high number of PI-RADS or Likert 3 lesions on fast MRI. These are classified as indeterminate or equivocal results and represent a dilemma for prostate MRI which will be addressed in subsequent chapters. There was low interobserver agreement for these indeterminate lesions, and this could lead to substantial variations in performance of MRI if incorporated into the screen-positive definition.

The high positive rate for MRI Score 3 further suggests that including these lesions may not represent the optimal threshold for fast MRI. The false positive rate is a major source of harm in prostate cancer screening due to the harmful effect of prostate biopsy and risk of overdiagnosis of insignificant prostate cancer. In contrast for MRI (Score 4-5), the false positive rate was 6.7% which provides a more acceptable balance of harms compared to MRI Score 3-5). In addition, the reproducibility at the higher threshold level was better than including score 3 so there will be less variation in performance between radiologists.

The comparison of detection rates suggests that fast MRI may be able to detect more clinically significant cancers than PSA at both thresholds. The detection rates were higher for threshold 3-5 for combined biopsy and for threshold 4-5 when results where stratified by biopsy detection method. Notably there was minimal overlap between the significant cancers detected by fast MRI and PSA given that only 35% of cancers were detected by both tests. This finding provides support for the hypothesis evaluated in Chapter 9 that a combination of PSA and MRI may be synergistic rather than using either test alone.

The key strength of my study was its pragmatism in evaluating a fast MRI protocol which has a fast scanning time and does not require intravenous contrast. The study was completed on 1.5T or 3.0T machines, without an endorectal coil and in both an academic and non-academic institution. We believed that this would make the results of IP1-PROSTAGRAM more reproducible and the technique deliverable in a high-volume screening setting. The conduct and reporting of PSA and fast MRI was conducted blind to other test results which minimized reporter bias. Radiologists were not present during the biopsy which ensured the reference test remained independent of the index. In addition, participants were blinded to the indication for biopsy until study completion so there was no selective withdrawal for each screening test. Despite the blinding a high rate of adherence to the MRI screening and recommendation to biopsy contributed to the strength of IP1-PROSTAGRAM. Finally, as a paired cohort study, the indication for biopsy was based on either of the tests under evaluation, blinded to the outcome of the other, allowing a robust comparison without any

incorporation bias. Last, by setting the threshold for biopsy based on MRI at 3 or above, I was able to evaluate whether a threshold of 3 or 4 was the more appropriate.

6.5.2 Comparison with previous studies

The primary outcomes of IP1-PROSTAGRAM show that MRI lesions (score 3-5) are detected in 1 in 6 men aged 50-69 years who undergo fast MRI screening for prostate cancer. The proportion of men who would have been recalled and had a false positive on biopsy due to a screen-positive MRI (score 3-5) was higher than for PSA. This was due to the fact that score 3 (equivocal) scans can be a normal finding in men in their early 50s due to diffuse changes in the peripheral zone. A similar finding was made in a pilot study evaluating a multi-parametric MRI in the general population, but the small sample size limited the study's ability to draw conclusions related to false positive findings and cancer detection rates[263].

In terms of the PSA findings, in this study the PSA threshold ≥ 3ng/ml had a high number of false negatives (missed significant cancer) as a result of the additional clinically significant cancers detected by fast MRI. The number of missed cancers was higher than expected compared to the Prostate Cancer Prevention (PCPT) trial. This could be for a number of reasons including verification bias which is addressed in the following chapter. Alternatively, the PCPT trial was conducted without modern imaging techniques and used 6-core sextant biopsy as a reference test so may have missed more significant cancers than the current 12 core systematic approaches.

Similar to PCPT, we found that the probability of PSA missing significant cancer was determined by the level of the PSA. The probability of significant cancer was present across a wide range of PSA values which suggests that the dichotomous division of PSA into 'normal' and 'abnormal' based on a level 3ng/ml is incorrect. PSA is a continuous measurement with decreasing probability at lower values.

In the present study, we noted that only three men with significant cancers missed by PSA had a PSA level below 1ng/ml. This supports previous studies which indicated there is minimal probability of clinically significant cancer at very low PSA levels. Given that there remains considerable diagnostic uncertainty beyond 1ng/ml, a subsequent chapter will evaluate whether PSA and fast MRI can be combined to optimise diagnostic outcomes in a multi-modal screening pathway.

6.5.3 Implications of findings

This chapter indicates that a threshold of 4 or greater may provide a more suitable balance between the benefits and harm for a fast MRI as a screening test. The proportion of men who would need to be recalled was similar to PSA as the number of clinically significant cancers detected was high. This was particularly apparent for the definition of significant cancer which incorporates both tumour volume and ISUP grade. Tumour volume is known to be associated with abnormal findings on MRI and has been shown to be predictive of tumour progression[264]. In contrast, a threshold of 3 or greater had clear disadvantages in terms of a higher false positive rate compared to PSA. The results highlight that the management of equivocal MRI scans (score 3) remains a clinical dilemma and solutions will be considered in subsequent chapters.

6.5.4 Limitations

This study has some limitations: First, although the paired screen-positive design is routine for screening studies, we acknowledge that not all participants received the reference test. This is required due to the ethical concerns of performing a prostate biopsy in healthy men from the general population without a clinical indication. The lack of a reference test meant that evaluation of an appropriate threshold for fast MRI could not integrate diagnostic accuracy measures such as the Youden Index. Further work is presented in Chapter 7 which corrects for the verification bias inherent in this study design.

Second, the supplementary community-based recruitment strategy to increase participation of ethnic minorities could introduce selection bias in recruitment. This strategy was chosen in response to previous PSA screening studies in which GP invitations failed to recruit ethnic minorities which limited the applicability of the results to a representative population. The response rate to the GP invitations for this screening trial was 16.8% which may represent further selection bias within the study population.

The comparability of our study to the general population is supported by the proportion of screen-positive PSA (10.0%) which is similar to the rate (10.6%) in a national cluster randomised trial which recruited 415,357 men. However, it is possible that the recruitment strategy could have changed the risk profile of the IP1-PROSTAGRAM participants. The prevalence of clinically significant prostate cancer was 4.2% (17 of 408) which is higher than previous screening studies such as PCPT. The reason for this difference could be explained by

a number of reasons beyond a higher risk population. For example, the changes in the Gleason grading system have led to a higher proportion of prostate cancers being classified as Gleason 3+4 so one would expect a higher prevalence of clinically significant prostate cancers following the latest system. In addition, the image-fusion targeted biopsy is more accurate as a reference standard than previous clinical trials such as PCPT which used TRUS biopsy.

Third, the modified PI-RADSv2 scoring system was developed for the diagnostic population and has not been validated for use in the general population. Radiologists were highly experienced in MRI interpretation in men with a raised PSA, but not experienced in using this technique in a high-volume screening setting. This could have contributed to the lower interobserver agreement compared to previous diagnostic accuracy studies. Interobserver agreement was variable depending on measures of agreement. When stratified by PI-RADS score, it was PI-RADS 3 (indeterminate lesions) which had the higher variability across each measure of interobserver variability. Further work is required to explore ways of optimising interpretation of prostate MRI for a screening population. This may require the development of a fast MRI scoring system for a screening population.

Due to the age of the cohort in IP1-PROSTAGRAM there was a high prevalence of diffuse peripheral zone changes and the PI-RADSv2 criteria do not provide clear guidelines for interpretation. Future studies should consider incorporating an additional standardization process prior to the start of the study to agree a consistent approach for reporting these cases.

Interobserver agreement was fair/low on certain measures such as kappa. It should be acknowledged that IP1-PROSTAGRAM was the first attempt for radiologists reporting prostate MRI without secondary clinical information such as PSA and it is possible disagreement could be due to inexperience with reporting a screening population. This may have been enhanced as PI-RADS may not be the most appropriate scoring method for screening as it has been designed for a secondary care (high-prevalence) population. There may have also been an influence of the prevalence of positive tests given that kappa showed fair agreement (values 0.2-0.4) while Gwet's AC1 coeffecient had higher levels of agreement for MRI threshold >/4 (>0.5).

Kappa is known to be affected by prevalence and even when agreement is high as occurred in IP1-PROSTAGRAM, kappa can be paradoxically low in a low prevalence population. We

attempted to control for this statistical paradox by stratifying the randomisation of scans selected for double reporting. However, this may not have been sufficient to correct for the prevalence bias particularly in the score 4-5 threshold. Other measures of interobserver agreement, such as Gwet's AC1, which are less affected by the prevalence paradox have higher levels of agreement. In the primary results paper, the statistical analysis plan selected overall agreement and kappa as the most common measures for interobserver agreement. However, using measures which are not affected by skewed distributions of categories such as the AC1 coefficient, leads to a higher agreement of 0.518 (moderate) for MRI score 4-5.

In addition, operators were not blinded to the indication for biopsy due to the use of image-fusion software which highlighted the MRI suspicious areas. As the reference test was not blinded there is a risk that operators could have been biased in favour of a particular screening test. This has been the standard methodology for the majority of trials in prostate cancer diagnostics which incorporate image-targeted biopsy into the reference standard[128]. To further minimize bias in this trial, all operators were required to follow the pre-contoured map which was prepared by another clinician prior to the biopsy procedure.

Finally due to the low prevalence of cancers in the general population, the study was not powered to detect differences in detection rates between each test. This is the reason that a statistical comparison of detection rates was not pre-specified in the statistical analysis plan.

6.6 Conclusion

In this prospective, population-based, cohort study, we found that MRI score 4-5 offers the most acceptable balance between risks and harms. At this score there is evidence that a fast MRI leads to increased detection of clinically significant prostate cancer compared to PSA without increasing the rate of overdiagnosis or number of prostate biopsies. If an MRI Score 3 (indeterminate) is included in the threshold, then the false positive rate reaches levels which would be unacceptable for a first-line screening test. Further work is required to determine the optimal management strategy for indeterminate lesions and this is explored in Chapter 11.

Chapter 7 – Addressing verification bias in IP1-PROSTAGRAM: Estimating sensitivity and specificity

7.1 Overview

The IP1-PROSTAGRAM trial used a paired screen-positive design (PSP), also known as a verification of only positive testers (VOPT) design. It is an inherent limitation with this design that not all participants receive the reference standard. In this chapter, I explore various correction methods to control for the verification bias within the IP1-PROSTAGRAM trial.

7.2 Introduction

7.2.1 Conceptual Framework 1: The reference standard

In order to evaluate the accuracy of a diagnostic test, the true disease status for each patient should be independently determined. The procedure to establish the disease status of a patient is known as the reference standard or gold standard. This should separate the target populations into two groups based on absence or presence of disease. To illustrate this, diagnostic accuracy studies use a classical two-by-two contingency table to separate diagnostic test result outcome (T) and True Disease Status (D):

Table 29: Standard two-by-two table for diagnostic accuracy studies

Test Result	True Disease status	
	D = 1	D =0
T = 1	TP	FP
T = 0	FN	TN

where TP: True positive, FP: False positive, FN: False negatives, TN True negative.

Sensitivity = TP/(TP+FN), Specificity = TN/(FN+TN), PPV = TP / (TP+FP), NPV = TN/(FN+FN)

The reference standard can be based on different approaches such as clinical assessment, biopsy or surgery but a key assumption of diagnostic accuracy studies is that the reference standard of disease accurately reflects the true disease status. Theoretically, the best available reference test for prostate cancer would be a pathological review of the whole prostate gland following removal via a radical prostatectomy. It is clear that this is not feasible nor ethical to perform such a procedure on a person without evidence of significant

cancer. The alternative diagnostic modality to be utilised as a reference standard has been prostate biopsy. Early diagnostic accuracy trials utilised TRUS biopsy but studies such as PROMIS have shown that this is inaccurate as a reference standard[88].

Therefore, various modifications to prostate biopsy techniques have been undertaken to increase the accuracy of this reference test. Prior to the validation of mpMRI, transperineal template-guided mapping biopsy was commonly utilised as the most accurate reference standard. This is an extensive biopsy technique which samples the gland at 5mm intervals via a brachytherapy grid fixed on a stepper and was the method used in the PICTURE study reported in Chapter 4. The technique is still subject to a small error but it remains the best available reference standard particularly if it incorporates targeted biopsy.

The disadvantage is that template mapping biopsy has a high adverse event rate which can include acute urinary retention in up to 24%, rectal pain in 26% and perineal pain in 41%[265]. This meant it was not feasible as a reference standard in IP1-PROSTAGRAM due to ethical reasons and the need to minimize harm in a population of healthy men who had been invited for screening. In IP1-PROSTAGRAM, a verification test which combined targeted biopsy with systematic biopsy was selected as the optimal reference standard. This was chosen as it is a technique which reduces toxicity while minimising mis-classification errors.

7.2.2 Conceptual Framework 2: Verification Bias

Even with the presence of an accurate reference test, it still may not be ethical to apply the test to all participants. This occurs when the reference test is invasive and/or expensive meaning that participants with screen-negative results may not receive it. When this occurs it is known as verification bias, also called ascertainment bias or work-up bias, where the determination of the true disease status depends on the result of the screening test rather than an independent reference standard applied to all participants.

Two types of verification bias are described in the literature:

1. **Partial verification bias**: This occurs when not all participants are subjected to the reference test. In this scenario, screen-positive tests are more likely to be verified than screen-negative results. This bias was evidenced in early studies evaluating the performance of PSA where only men with an abnormal PSA or abnormal digital rectal examination (DRE) had the reference standard. The impact of this bias on sensitivity and specificity depended on how the researchers dealt with the unverified cases.

In Morgan et al (1996)[266], the men with a screen-negative PSA and DRE test were presumed to not have prostate cancer and were included as true negatives (Table 30). This approach will lead to an under-representation of false negatives and overestimation of true negatives given not all the screen-negatives will be disease-free. This is represented in the two-by-two table and leads to a biased estimate of diagnostic accuracy with both sensitivity and specificity overestimated.

Table 30: Verification bias when unverified cases are assumed to be disease free (D=0).

Test Result	True Disease status (V =1)	
	D = 1	D =0
T = 1	TP	FP
T = 0	FN ↓	TN ↑
	Sensitivity ↑	Specificity ↑

This approach shifts the balance of true and false negatives in favour of true negative leading to an overestimation of both sensitivity and specificity

One solution to address this bias is to design a study such that the reference test is applied to a random sample of screen-negative participants for disease verification. The sample size of this sub-group would need to be sufficient to generate an adequate number of false negatives to estimate sensitivity. This design would have been challenging in IP1-PROSTAGRAM where the prevalence of clinically significant cancer was estimated at around 2%. In this scenario, 1,000 men would need to be verified in the screen-negative arm to generate 20 false negative results.

Although this strategy would address verification bias it would be at a significant cost; both to participants undergoing an invasive biopsy without clinical indication and the research costs associated with an expensive verification test. In addition, previous attempts to follow the random screen-negative verification approach in cancers outside prostate cancer have not been successful as participants who were screen-negative were often not compliant with attending the reference test resulting in further biasing the results[267].

Due to these reasons this approach was not used in IP1-PROSTAGRAM which instead adopted a PSP design. This design is a recognized approach in the literature and has been used in other screening studies for prostate cancer[249]; notably this was used by the Stockholm-3 (STHLM3) novel biomarker panel to report the ROC curves and AUC values for PSA vs. STHML3.

2. **Differential verification bias**: This occurs when researchers attempt to address partial bias by applying a less accurate reference standard to the screen-negative group while the accurate reference standard is reserved for screen-positive participants. If the inaccurate reference standard fails to identify false-negative results, then the diagnostic performance of the test will be overestimated.

The use of a composite reference standard has been common in breast cancer screening trials into mammography and/or breast MRI where screen-negative participants have longitudinal follow-up with clinical assessments while screen-positive participants have a targeted biopsy. In Berg et al (2012)[268] the sensitivity/specificity of MRI screening for breast cancer was reported despite the biopsy reference standard being applied only to 8.5% of the MRI group. In this study, one year follow up was incorporated into the reference standard to establish the true vs. false negative rate.

This approach was not replicated in IP1-PROSTAGRAM as clinical follow-up is a poor reference standard for prostate cancer. Due to the long natural history of our disease this method of verification would only yield results over 5 to 15 years. In the short term using this approach would have only artificially increased the number of true negatives without providing useful information on false negatives. Similar to Table 30 where participants are assumed to be screen-negative, the impact is inflated sensitivity and specificity results.

7.2.3 Aims and objectives

It is clear that evaluating a test in a screening population requires a different conceptual framework to the classical view that a reference test must be applied in all. Indeed, in situations where the reference standard cannot be obtained there are recognised methods in the literature to correct for verification bias. The aim of this chapter is to explore the different methods of correcting for verification bias and present point estimates for sensitivity and specificity for each MRI threshold from the IP1-PROSTAGRAM trial.

7.3 Methods

This chapter utilized the data from participants of the IP1-PROSTAGRAM study. The trial procedures have been described in Chapters 5-6. In brief, 408 men underwent prostate cancer screening with PSA and new imaging tests which included a fast MRI and shear wave elastrography (SWE). A biopsy was recommended in the presence of any screen-positive test. The biopsy was a transperineal systematic 12-core biopsy with or without MRI-targeted biopsy of any areas on imaging suggestive of prostate cancer. Herein, we present several methods which have been described to correct for verification bias.

7.3.1 Complete Cases Analysis

In a complete case analysis, the diagnostic calculations are performed using the standard formulae but only using the participants who underwent verification[269]. This method was used by Catalona et al (1994)[270] in which screen-negative participants were excluded from the analysis if they had a PSA less than 4ng/ml and a normal DRE, so did not undergo a reference test. Although this approach has been widely used in screening studies it can result in an overestimation of sensitivity and underestimation of specificity as has been illustrated in this two by two table.

Table 31: Complete cases analysis

Test Result	True Disease status (V =1)	
	D = 1	D =0
T = 1	TP	FP
T = 0	FN ↓	TN ↓
	Sensitivity ↑	Specificity ↓

Effect when unverified cases are removed from the analysis. This approach removes both false and true negatives and leads to an overestimation of sensitivity and underestimation of specification

7.3.2 Composite reference standard with exclusion of the index test

The alternative to a complete cases analysis is to combine a series of tests to construct a composite reference standard. Composite reference standards are appealing since combining several tests for screen-negative patients may provide a more accurate

perspective on disease status than simply excluding screen-negative patients. By excluding the index test from the reference standard one can avoid incorporate bias[271].

In theory this approach would be possible in IP1-PROSTAGRAM as multiple diagnostic tests were performed during the screening evaluation. These included a PI-RADS score, a Likert score, an ultrasound score and a PSA test. Using this approach the composite reference standard would be the following for a screen-negative index test:

a. Screen-negative PSA: MRI (PI-RADS or Likert) and SWE

b. Screen-negative MRI: PSA and SWE

This approach is based on combining multiple (presumed) imperfect reference standards. This leads to a risk of persistent misclassification of false negatives. Such misclassification could underestimate false negatives and lead to an overestimation of both sensitivity and specificity as shown in this 2 x 2 table (Table 32):

Table 32: Composite reference standard

Test Result	True Disease status (V =1)	
	D = 1	D =0
T = 1	TP	FP
T = 0	FN ↓	TN ↑
	Sensitivity ↑	Specificity ↑

Effect of composite reference standard with multiple imperfect reference standard. This approach can lead to an overestimation of both sensitivity and specificity

7.3.3 Multiple Imputation

An alternative approach is to consider partial verification as a missing data problem and employ multiple imputation methods originally described by Harel and Zhou[272]. This method involves replacing the missing reference value by a reasonable estimated value based on an estimate from the specific subgroup to which the subject belongs. The process is repeated multiple times to represent the uncertainty associated with this method. The imputed datasets are combined to provide an overall mean score.

To impute the missing verification levels, this chapter uses the MICE (Multivariate Imputation via Chained Equations) package in R statistical software. This procedure assumes that the

distribution of the missing verification values can be modelled on the basis of the results from the other tests in combination with the outcome of the known verified test. This procedure was repeated 1,000 times to ensure convergence of the mean results. Multiple imputation methods have been shown to provide superior results in comparison to complete case analysis and composite reference standard in simulation studies.

7.3.4 Begg and Greene (B&G) Correction Method:

The final method which I will use is the B&G method[273] which is one of the most common statistical approaches to correct for verification bias. It uses an inverse probability weighting to establish an estimate of diagnostic accuracy. It assumes that the verification bias is solely the result of the preferential referral of patients with screen-positive tests for biopsy. This assumption is met in IP1-PROSTAGRAM as the patients selected for verification depended only on the results of the screening tests and was independent of other variables.

This method has been widely used in previous diagnostic accuracy studies[274-276]. For example Punglia et al[275] (2003) evaluated the performance of PSA in 6,601 men who underwent screening for prostate cancer. Verification by biopsy was determined by both pre-test referral characteristics and a screen-positive PSA. When this was performed for PSA, the overall performance of PSA was reduced so the authors recommended a lower threshold value for PSA to improve the diagnostic performance of PSA.

The correction method involves calculating the observed proportion of disease and non-disease in the participants who underwent the reference test. It then calculates the expected number of diseased and non-diseased among the nonverified patients equivalent to inflating the two-by-two table by multiplying each cell with the inverse probability of having been verified by the reference standard.

The classical two-by-two table is updated to;

Screening Test	Disease (D) Verified (V =1)		Disease not verified (V=0)	Total
	D = 1	D =0		
$T = 1$	α_1	b_1	u_1	n_1
$T = 0$	α_2	b_2	u_2	n_2

Adjusted sensitivity and specificity can then be calculated using the following formulae:

$$Adjusted\ Sensitivity\ [P(T1\ |\ D1)] = \frac{n1\alpha1\ /\ (\alpha1\ +\ b1)}{n1\alpha1\ /\ (\alpha1\ +\ b1)\ +\ n2\alpha2\ /\ (\alpha2\ +\ b2)}$$

$$Adjusted\ \text{Specificity}\ [P(b2\ |\ D0)] = \frac{n2b2\ /\ (\alpha2\ +\ b2)}{n1b1\ /\ (\alpha1\ +\ b1)\ +\ n2b2\ /\ (\alpha2\ +\ b2)}$$

7.3.5 Outcomes

The primary outcomes were a comparison of sensitivity and specificity using each correction method. This was performed across two definitions of clinical significance; any Gleason ≥ 3+4 (primary definition) and Gleason ≥4+3 and/or MCCL ≥6mm of any grade. Secondary outcomes included reporting the point estimates for sensitivity, specificity, PPV and NPV across each screen-test using multiple definitions of significant cancer.

7.3.6 Statistical methods

All analyses for ITS were conducted in R Version 4.0.2[213] using R Studio Version 1.0.44 and the MICE package for imputation. Differences between sensitivity and specificity were compared using different tests depending on the correction method. The differences between complete cases analysis and composite reference standards were tested by use of McNemar's. Due to the correction method, the overlap of participants was available between each test which allowed computation of the McNemar. For the B&G method, a McNemar could not be performed as the overlap was not known. Therefore, we compared the outcomes from the B&G method using a z test which was performed on the differences between sensitivity and specificity.

7.4 Results

7.4.1 Primary outcomes

Figure 35 summarises the estimates of sensitivity and specificity for each test and correction method with 95% confidence intervals. Each correction method showed the same trend when it came to point estimates of sensitivity, with MRI Score ≥ 3 highest, followed by MRI Score ≥ 4, with PSA found to be lowest. There was some variation in the actual estimates for sensitivity depending on correction method. The point estimate for sensitivity of MRI score

≥ 3 was 61.2% and 82.4%, MRI Score ≥ 4 was between 42.4% and 64.7% while PSA was between 22.1% and 41.2%.

Although the point estimates for sensitivity were variable, the differences between each test were consistent despite the correction method. The comparisons were consistent in showing that fast MRI was more sensitive than PSA for Gleason ≥ 3+4 at a threshold score ≥ 3 (difference 39.1%-41.2%, p≤0.02). The sensitivity for MRI Score ≥ 4 was between +20.3% to +23.5% higher than PSA indicating a trend towards higher sensitivity, although the results failed to reach statistical significance (p=0.077, B&G method). This can be seen in Figure 35 where the confidence intervals for the sensitivity are wide due to the low prevalence of significant disease. This has prevented us drawing firm conclusions related to MRI Score ≥ 4.

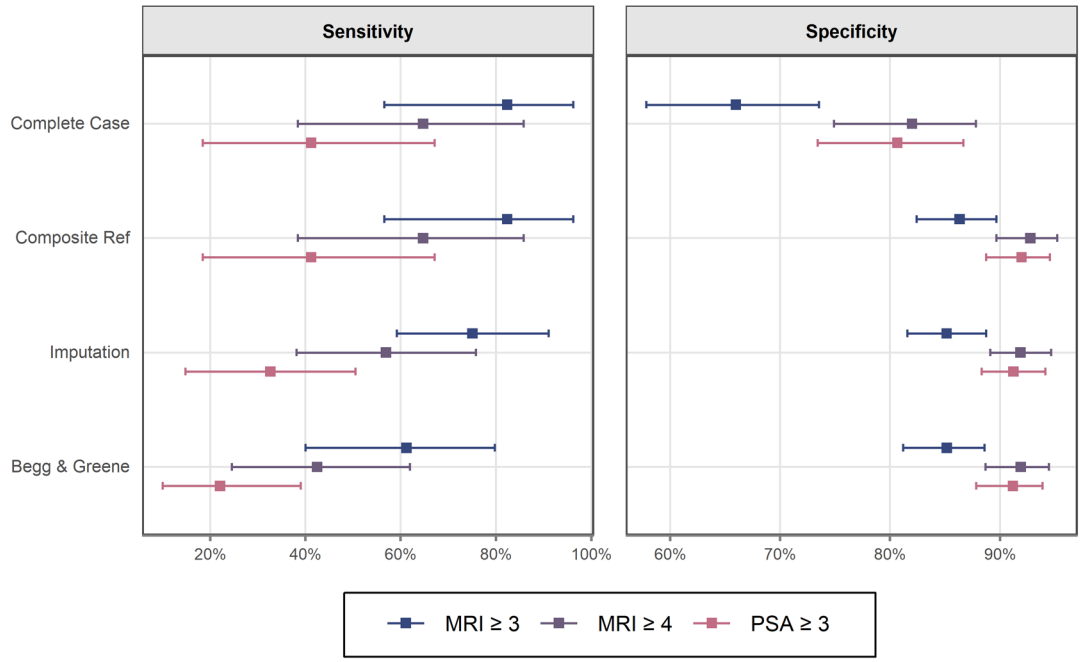


Figure 35: Comparison of sensitivity and specificity for each correction method for detection of Gleason ≥ 3+4.

In terms of specificity, the point estimates were less variable apart from the complete case correction method. All correction methods found a similar trend that MRI Score ≥ 3 had the lowest sensitivity while there was no difference in sensitivity between MRI Score 4 and PSA. The point estimates for specificity of MRI score ≥ 3 was between 66.0% and 85.2%, MRI Score ≥ 4 was between 82.0% and 92.8% while PSA was between 80.7% and 92.0%. The comparisons between tests confirmed that MRI score ≥ 3 was less specific than PSA (difference -6.1-14.7, p ≤ 0.02) while there was no difference in specificity between MRI Score ≥ 4 and PSA (difference 0.7-1.3%, p=0.674-0.722).

A similar comparison was performed for secondary definition of significant disease (Gleason ≥4+3 and/or MCCL ≥6mm. Due to the lower prevalence of this disease the confidence intervals were wider. The point estimates for sensitivity appeared to show less variation between MRI Score ≥ 3 and MRI Score ≥ 4. The outcomes for specificity remained similar to the results for any Gleason ≥ 3+4.

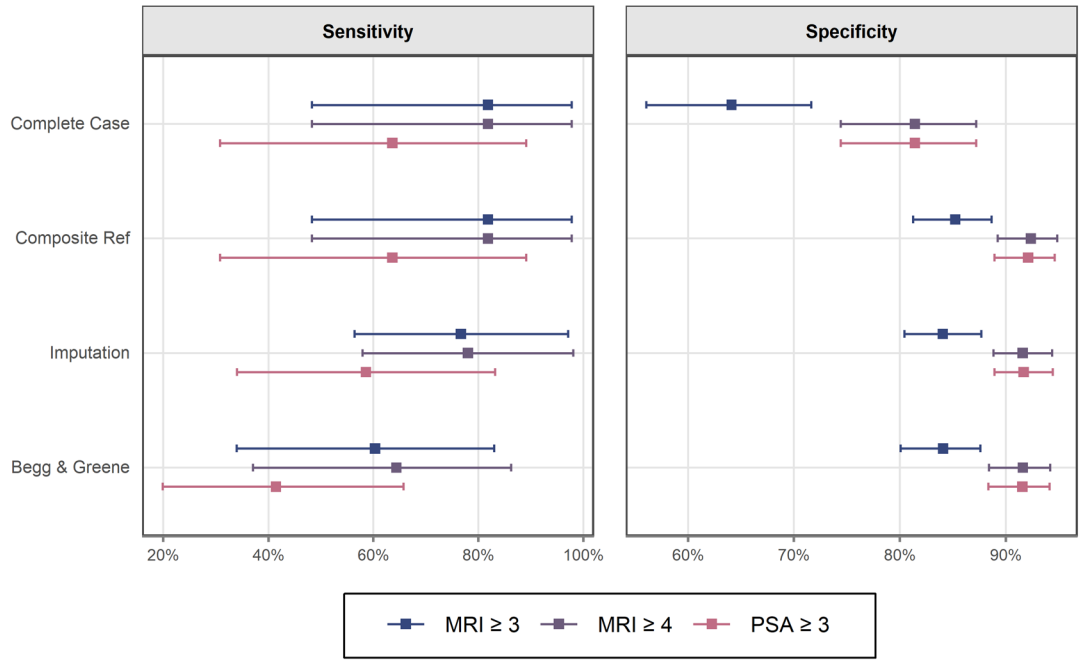


Figure 36: Comparison of sensitivity and specificity for each correction method at the secondary definition.

7.4.2 Secondary Outcomes

7.4.2.1 Complete cases analysis

In total, 167 participants had a biopsy and so had data for the complete case analysis. Point estimates of each screening test's sensitivity and specificity for detection of Gleason ≥ 3+4 were 82.4% (95% CI 56.6-96.2) and 66% (95% CI 57.8-73.5) for MRI Score ≥3, 64.7% (95% CI 38.3-85.8) and 82% (95% CI 74.9-87.8) for MRI Score ≥4 and 41.2% (95% CI 18.4-67.1) and 80.7% (95% CI 73.4-86.7) for PSA. Additional definitions are shown in Table 33.

Table 33: Diagnostic accuracy of MRI and PSA using complete case analysis

	MRI Score ≥ 3	MRI Score ≥ 4	PSA ≥ 3 ng/ml	p-values*	
				MRI ≥3 vs PSA	MRI ≥4 vs PSA
Any Gleason ≥3+4 (Prevalence 10.2%)					
Sensitivity	82.4% (56.6% to 96.2%)	64.7% (38.3% to 85.8%)	41.2% (18.4% to 67.1%)	0.02	0.102
Specificity	66.0% (57.8% to 73.5%)	82.0% (74.9% to 87.8%)	80.7% (73.4% to 86.7%)	0.011	0.777

Table 33: Diagnostic accuracy of MRI and PSA using complete case analysis					
	MRI Score ≥ 3	MRI Score ≥ 4	PSA ≥ 3 ng/ml	p-values*	
				MRI ≥3 vs PSA	MRI ≥4 vs PSA
UCL/Ahmed 1: Gleason ≥ 4+3 and/or MCCL ≥ 6mm (Prevalence 6.6%)					
Sensitivity	81.8% (48.2% to 97.7%)	81.8% (48.2% to 97.7%)	63.6% (30.8% to 89.1%)	0.317	0.317
Specificity	64.1% (56% to 71.6%)	81.4% (74.4% to 87.2%)	81.4% (74.4% to 87.2%)	0.002	1.00
UCL/Ahmed 2: Gleason ≥ 3+4 and/or MCCL ≥ 4mm (Prevalence 12%)					
Sensitivity	70.0% (45.7% to 88.1%)	55.0% (31.5% to 76.9%)	50.0% (27.2% to 72.8%)	0.248	0.739
Specificity	65.3% (57% to 73%)	81.6% (74.4% to 87.5%)	82.3% (75.2% to 88.1%)	0.003	0.884

* McNemar tests to compare sensitivity and specificity

7.4.2.2 Composite reference standard

There were 390 participants eligible for the composite standard method. Of 408 participants, two did not complete the fast MRI and 16 withdrew despite a recommendation for biopsy. This method does not affect the level of true positives, false positives or false negatives so sensitivity and PPV remained the same as the complete method. Due to the increase in true negatives there were differences in the estimates for specificity and NPV. The complete results for the composite reference standard are shown in Table 34.

Table 34: Diagnostic accuracy of MRI and PSA using a composite reference standard					
	MRI Score ≥ 3	MRI Score ≥ 4	PSA ≥ 3 ng/ml	p-values*	
				MRI ≥3 vs PSA	MRI ≥4 vs PSA
Any Gleason ≥3+4 (Prevalence 10.2%)					
Sensitivity	82.4% (56.6% to 96.2%)	64.7% (38.3% to 85.8%)	41.2% (18.4% to 67.1%)	0.02	0.102
Specificity	86.3% (82.4% to 89.6%)	92.8% (89.6% to 95.2%)	92.0% (88.7% to 94.5%)	0.015	0.674
UCL/Ahmed 1: Gleason ≥ 4+3 and/or MCCL ≥ 6mm (Prevalence 6.6%)					
Sensitivity	81.8% (48.2% to 97.7%)	81.8% (48.2% to 97.7%)	63.6% (30.8% to 89.1%)	0.317	0.317
Specificity	85.2% (81.2% to 88.6%)	92.3% (89.2% to 94.8%)	92.1% (88.9% to 94.6%)	0.004	1.00
UCL/Ahmed 2: Gleason ≥ 3+4 and/or MCCL ≥ 4mm (Prevalence 12%)					
Sensitivity	70.0% (45.7% to 88.1%)	55.0% (31.5% to 76.9%)	50.0% (27.2% to 72.8%)	0.248	0.891
Specificity	86.2% (82.3% to 89.6%)	92.7% (89.6% to 95.1%)	92.7% (89.6% to 95.1%)	0.005	1.00

* McNemar tests to compare sensitivity and specificity

7.4.2.3 Multiple Imputation

There were 406 participants who received both fast MRI and PSA screening tests so were eligible for the multiple imputation. An example of the outcomes from 10 imputed datasets using the MICE statistical approach for multiple imputation is shown in Table 35. The full imputation included 1,000 datasets.

Table 35: 10 out of 1,000 imputed datasets for MRI≥3 & Gleason ≥3+4

		T = 1	T = 0	Sensitivity	Specificity
Data Set 1	D = 1	15	57	0.83	0.85
	D = 0	3	331		
Data Set 2	D = 1	14	58	0.82	0.85
	D = 0	3	331		
Data Set 3	D = 1	17	55	0.85	0.86
	D = 0	3	331		
Data Set 4	D = 1	14	58	0.82	0.85
	D = 0	3	331		
Data Set 5	D = 1	14	58	0.82	0.85
	D = 0	3	331		
Data Set 6	D = 1	19	53	0.86	0.86
	D = 0	3	331		
Data Set 7	D = 1	14	58	0.82	0.85
	D = 0	3	331		
Data Set 8	D = 1	14	58	0.22	0.83
	D = 0	51	283		
Data Set 9	D = 1	14	58	0.82	0.85
	D = 0	3	331		
Data Set 10	D = 1	14	58	0.82	0.85
	D = 0	3	331		

D = 1 where disease present; D = 0 where no disease present

T = 1 for positive test; T=0 for negative test result

When the 1,000 imputed datasets were combined the prevalence of significant cancer was lower across all definitions. Any Gleason ≥ 3+4 was 4.9%, UCL/Ahmed 1 was 2.95% and UCL/Ahmed 2 was 5.2%. In terms of diagnostic accuracy, point estimates for sensitivity of Gleason ≥ 3+4 were 75.1% (95% CI 59.1-91) and 85.1% (95% CI 81.6-88.7) MRI Score ≥3, 56.9% (95% CI 38.1-75.8) and 91.9% (95% CI 89.1-94.6) for MRI Score ≥4 and 32.6% (95% CI 14.8-50.5) and 91.2% (95% CI 88.3-94.1) for PSA. Other definitions are shown in Table 36.

Table 36: Diagnostic accuracy of MRI and PSA using multiple imputation					
	MRI Score ≥ 3	MRI Score ≥ 4	PSA ≥ 3 ng/ml	p-values*	
				MRI ≥3 vs PSA	MRI ≥4 vs PSA
Any Gleason ≥3+4 (Prevalence 4.9%)					
Sensitivity	75.1% (59.1% to 91%)	56.9% (38.1% to 75.8%)	32.6% (14.8% to 50.5%)	-	-
Specificity	85.1% (81.6% to 88.7%)	91.9% (89.1% to 94.6%)	91.2% (88.3% to 94.1%)	-	-
UCL/Ahmed 1: Gleason ≥ 4+3 and/or MCCL ≥ 6mm (Prevalence 2.95%)					
Sensitivity	76.7% (56.3% to 97%)	78% (58% to 98.1%)	58.6% (34% to 83.2%)	-	-
Specificity	84% (80.4% to 87.7%)	91.6% (88.8% to 94.3%)	91.7% (88.9% to 94.4%)	-	-
UCL/Ahmed 2: Gleason ≥ 3+4 and/or MCCL ≥ 4mm (Prevalence 5.2%)					
Sensitivity	57.9% (41.6% to 74.3%)	43.8% (27.1% to 60.5%)	40.3% (23.3% to 57.3%)	-	-
Specificity	84.7% (81% to 88.4%)	91.6% (88.7% to 94.4%)	92.1% (89.3% to 94.8%)	-	-

* P values cannot be calculated due to imputation methodology

7.4.2.4 Begg and Greene Correction Method

The B&G method utilised the 406 participants which was the same group as used by multiple imputation. Using this method, we estimated sensitivity and specificity for detection of Gleason ≥ 3+4 as 61.2% (95% CI 40-79.7) and 85.2% (95% CI 81.2-88.6) for MRI Score ≥3, 42.4% (95% CI 24.5-62) and 91.9% (95% CI 88.7-94.4) for MRI Score ≥4, and 22.1% (95% CI 10-38.9) and 91.2% (95% CI 87.8-93.8) for PSA. A summary of results for this final correction method is shown in Table 37.

Table 37: Diagnostic accuracy of MRI and PSA using the Begg and Greene correction method					
	MRI Score ≥ 3	MRI Score ≥ 4	PSA ≥ 3 ng/ml	p-values*	
				MRI ≥3 vs PSA	MRI ≥4 vs PSA
Any Gleason ≥3+4 (Prevalence 8.8%)					
Sensitivity	61.2% (40% to 79.7%)	42.4% (24.5% to 62%)	22.1% (10% to 38.9%)	0.002	0.077
Specificity	85.2% (81.2% to 88.6%)	91.9% (88.7% to 94.4%)	91.2% (87.8% to 93.8%)	0.011	0.722
UCL/Ahmed 1: Gleason ≥ 4+3 and/or MCCL ≥ 6mm (Prevalence 4.69%)					
Sensitivity	60.4% (33.9% to 83%)	64.4% (37% to 86.2%)	41.4% (19.9% to 65.7%)	0.259	0.176
Specificity	84.1% (80.1% to 87.6%)	91.6% (88.4% to 94.1%)	91.5% (88.3% to 94.1%)	0.001	0.983

Table 37: Diagnostic accuracy of MRI and PSA using the Begg and Greene correction method					
	MRI Score ≥ 3	MRI Score ≥ 4	PSA ≥ 3 ng/ml	p-values*	
				MRI ≥3 vs PSA	MRI ≥4 vs PSA
UCL/Ahmed 2: Gleason ≥ 3+4 and/or MCCL ≥ 4mm (Prevalence 9.66%)					
Sensitivity	44.1% (27.5% to 61.8%)	33% (18.6% to 50.1%)	28.8% (15.5% to 45.4%)	0.168	0.692
Specificity	84.8% (80.7% to 88.3%)	91.7% (88.4% to 94.3%)	92% (88.8% to 94.6%)	0.002	0.880

* Z tests were used to compare sensitivity and specificity

7.5 Discussion

7.5.1 Principle findings

When verification bias occurs by design the pattern of missing data is known and so can be adjusted for using well-established correction methods which have been described in this chapter. Using these methods, I found that MRI Score ≥ 3 was more sensitive, albeit less specific, than PSA for identification of any Gleason ≥ 3+4. The finding of a nonstatistical trend across all correction methods (p=0.07) in favour of MRI Score ≥ 4 having a higher sensitivity than PSA ≥ 3ng/ml (sensitivity difference +20.3 to +24.3%) is particularly interesting and may be explained by the lack of statistical power inherent to the relative small sample size for a screening study.

Each correction method produced different point estimates for sensitivity and specificity, but the trend was consistent with fast MRI having a high sensitivity. The specificity for MRI Score ≥ 3 was reduced in comparison to PSA, while MRI Score ≥ 4 was not statistically significantly different to PSA ≥ 3. The highest sensitivity for fast MRI and PSA was found by the complete case method which was also the method of correction that deviated the most compared to the other methods across all definitions of significant disease.

This can be explained by the methodology used in the complete case analysis where all non-verified cases are removed from the analysis, so the population has fewer men who were screen-negative. This approach estimated a sensitivity of 82.4%, 64.7% and 41.2% for MRI ≥ 3, MRI ≥ 4 and PSA, respectively. The method of removing participants without verification leads to persistent bias in the results and an inevitable underestimation of sensitivity and overestimation of specificity

In contrast, a composite reference standard assumes that the rate of false negatives in the screen-negative group is negligible. This method also has limitations as the screen-negative composite reference standard remains an imperfect reference test meaning that if both reference tests are screen-negative it is likely that residual misclassification will still be present. There are also issues related to differential verification bias as there is a different reference test for screen-negative and screen-positive patients.

Both the complete case method and composite reference standard have limitations so alternative statistical methods were used to estimate diagnostic accuracy. The Begg and Greene method estimates diagnostic accuracy by calculating the number who had significant cancer on verification and projects these estimates onto the corresponding men who had no verification. The Begg and Greene correction method produced sensitivity estimates of 61.2%, 42.4% and 22.1% for MRI ≥ 3, MRI ≥ 4 and PSA respectively. The estimates of specificity were 85.2%, 91.2% and 91.2%.

However, this method also has limitations as it assumes that those with a screen-negative test who did undergo verification due to a screen-positive from a different test would have a similar cancer detection rate to those that did not undergo verification as all screening tests were negative. In reality, this is unlikely to be the case in IP1-PROSTAGRAM as verification was based on multiple tests including a PI-RADS Score, Likert Score, Ultrasound Score and PSA level. It can be hypothesised that men with a screen-negative test of multiple modalities are likely to be at lower risk than those with a screen-positive alternative test but still negative on the index test.

The final method was conditional imputation of the missing verification data which has been described as superior to the B&G method[272, 277]. The B&G method can be considered as an example of single imputation where the disease status of the unverified participants is imputed based on the probability of cancer in the verified group with a screen-negative result. The reason for unverified data is more complex in IP1-PROSTAGRAM as it was not based on a single screening test. Instead the pattern of missing values is determined by multiple tests. Using this method, sensitivity was estimated as 75.1%, 56.9% and 32.6% while specificity was 85.1%, 91.9% and 91.2% for MRI ≥ 3, MRI ≥ 4 and PSA respectively.

7.5.2 Comparison with previous studies

7.5.2.1 Fast MRI Diagnostic Accuracy

There has been a single pilot study evaluating the performance of a standard multiparametric MRI in a screening population but the sample size was not sufficient to comment on diagnostic accuracy[119]. Due to the lack of data in a screening population, the only data available for comparison is estimates of sensitivity and specificity in a secondary care population where there have been multiple clinical trials.

Our estimated sensitivity of between 61.2% to 82.2% for fast MRI, at a threshold of 3 or greater, is lower than previous meta-analysis of published studies into MRI within diagnostic secondary care setting[278]. A Cochrane meta-analysis reported a pooled sensitivity of 89% (95%CI 0.82-0.94) for MRI score ≥3 and 72% (95% CI 0.52-0.86) for MRI score ≥4[128]. In addition, our estimated specificity of between 66.0% and 86.3% are higher than the Cochrane meta-analysis where the estimates for specificity were 39% (95% CI 0.32-0.47) for MRI Score ≥3 and 78% (95% CI 0.68-0.86) for MRI score ≥4[128].

In practice, the sensitivity and specificity of a screening test will vary across different clinical populations, a phenomenon referred to as the spectrum effect[150]. Traditionally, sensitivity and specificity were assumed to be fixed test characteristics; however it is increasingly recognised that sensitivity and specificity vary with disease prevalence. Several studies have reported that specificity improves in populations with lower prevalence of disease[279, 280]. Simulation studies have shown that a test developed on a sample with a higher disease prevalence will generally have a lower sensitivity and higher specificity when utilised in a population with lower disease prevalence[281, 282].

A visualisation of the impact of the spectrum effect in prostate cancer diagnosis is shown in Figure 37 where prevalence of prostate cancer has been changed due to variations in PSA distribution in screening, diagnostic and radical prostatectomy populations. The values used in this graph are for illustration only and the exact gradient of the curves could be different in real-world environments. The values were obtained via simulations within scenarios from Usher-Smith et al (2016)[282].

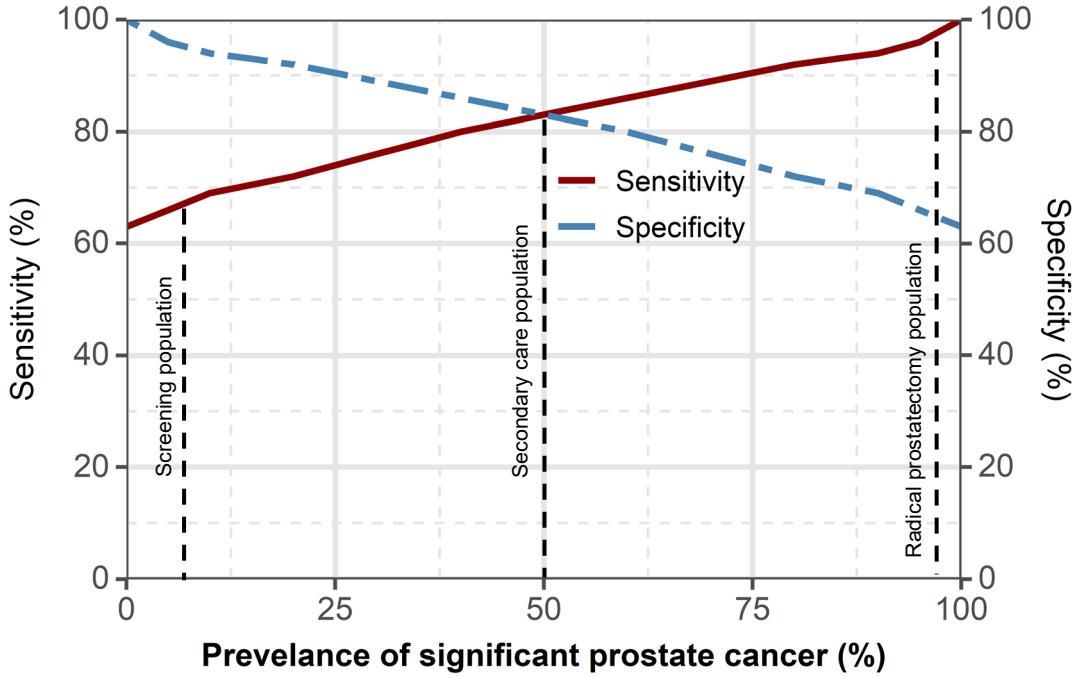


Figure 37: Simulated variation in sensitivity and specificity of MRI in different populations due to the spectrum effect. The graphs shows the variations in sensitivity (solid red line) and specificity (blue dashed line) with prevalence of significant prostate cancer where prevalence has been altered by varying the distribution of PSA values while keeping the threshold value of test constant. Values for plots obtained via simulation within scenario from Usher-Smith et al (2016) [282]. Clinically significant prostate cancer defined as Gleason >/= 3+4. The prevalence of significant prostate cancer in the screening population has been assumed at 5%, in a secondary population as 50% and radical prostatectomy population as 95% to account for cases of Gleason 3+3 disease.

7.5.2.2 Diagnostic Accuracy of PSA

The design of IP1-PROSTAGRAM also provides a valuable opportunity to re-evaluate the diagnostic accuracy of PSA, using a threshold of 3ng/ml, with contemporary prostate biopsy techniques and MRI-fusion software targeted biopsies. The early PSA trials provide historic evidence that a threshold of 3ng/ml will miss clinically-significant disease[158] and in the UK Cluster randomised trial of PSA testing for Prostate cancer (CAP)[283], 68 of 146 (46.6%) men dying from prostate cancer with a valid screening test had a PSA <3ng/ml.

Modern trials evaluating new biomarkers for screening have used a lower PSA threshold such as the STHLM3 trial which did not biopsy men with a PSA ≤ 1ng/ml[249]. This study reported an AUC of 0.56 (95% CI 0.54-0.59) for PSA compared to 0.74 (0.72-0.75) using the biomarker with DRE and prostate volume. As these trials have applied an a priori PSA threshold they cannot comment on performance of PSA across the full range of levels[42].

Historic evidence comes from the PCPT trial[158] which found a higher sensitivity of 57.6% for Gleason ≥ 3+4 in contrast to our estimate which varied between 22.1%-42.1% depending on the correction method. A possible explanation for this was that the PCPT trial was conducted

without modern imaging techniques and used 6-core sextant TRUS biopsy as a reference test. As a reference standard, TRUS is an imperfect test and would be expected to underestimate disease burden in men.

7.5.3 Implications of findings

In Chapter 3 (sub-section 3.4.4.1), I discussed that one of the potential problems of MRI was a low specificity as previous studies such as PROMIS have a found a specificity of 41% for MRI Score ≥ 3[88]. This level of specificity could have precluded consideration as a screening test due to excessive false positives. However, the correction methods of Begg and Greene and multiple imputation found specificities between 86.3%-92.8% for MRI Score ≥ 3 and 85.1% and 91.9% for MRI Score ≥ 3 which is encouraging.

It is important for fast MRI to have a high specificity for significant prostate cancer as even highly specific tests will lead to a high number of false-positive results when applied across a large population over many screening cycles. For example although the specificity of screening mammography is estimated between 94-97%[284, 285], the cumulative false positive rate is high and after 10 years of screening more than half of women in the USA receive at least one false-positive recall[286].

Due to the higher specificity of MRI Score ≥ 4 compared to ≥ 3, the findings of this chapter support the conclusion of Chapter 6 that the threshold of MRI Score 4 may provide a more suitable balance between the potential benefits and harms of screening. At this threshold, the specificity of fast MRI was maintained in comparison to PSA although the estimated specificity of MRI Score ≥ 4 remains lower than mammography due a number of benign entities which cause MRI signal abnormalities and yield false-positive results. These include acute/chronic prostatitis, glandular BPH, post-inflammatory glandular atrophy and fibrosis[180].

It is possible that further improvements in specificity of fast MRI may be seen across screening rounds as radiologists benefit from previous scans for comparison. The specificity of mammography has been shown to increase from 87% to 96% when previous mammograms were available, as radiologists were more confident in reporting equivocal lesions showing no change as benign[287].

In terms of sensitivity, we found that MRI Score ≥ 4 maintained a high performance level particularly for the definition of significant cancer which incorporates both tumour volume

and cancer grade. At this secondary definition of Gleason ≥ 4+3 and/or MCCL ≥ 6mm, the sensitivity of MRI Score ≥ 4 was between 64.4% to 81.8%. Our findings concur with findings in the diagnostic setting that certain small volume disease is difficult to visualise on MRI.

We acknowledge that there is a trade-off to be considered with an MRI threshold of 4 related to a reduction in the detection of low-volume Gleason 3+4 (ISUP 2) tumours, but with the benefit of higher specificity and a corresponding lower biopsy and false-positive rate. If an MRI score ≥ 4 was used as the threshold to denote a screen-positive MRI, 17.6% of men with Gleason ≥3+4 would not be recommended for a biopsy. Using a threshold MRI Score ≥ 3 would increase detection of low volume Gleason ≥3+4 but at a cost of excessive referral to biopsy.

7.5.4 Limitations

The analyses I have performed in this chapter have several limitations.

First, the IP1-PROSTAGRAM trial was not powered to detect differences in diagnostic accuracy between PSA and fast MRI. The primary objective of IP1-PROSTAGRAM was to evaluate the positive test rate of MRI and PSA so the sample size was calculated for this primary endpoint. A study evaluating diagnostic accuracy will require a much larger sample size due to the low prevalence of significant cancer in this population. Due to this low event rate, we cannot draw firm conclusions on the impact of fast MRI on sensitivity although there was a trend favouring MRI Score ≥ 4 compared to PSA.

Second, the study was conducted at two institutions which are recognized for their expertise in radiology and diagnosis and treatment of prostate cancer. It is possible that the results may not be replicated in other centres.

Third, we acknowledge that the reference standard was not conducted with operators blinded to the results of the impact test. This may have introduced bias into the reference standard which can occur when operators have a preconceived idea about the new test which subconsciously influences the performance of the reference test. This can be a particular problem for reference tests which are dependent on multiple samples taken for histological analysis. Previous studies have suggested that operators may vary the number of samples according to individual assessment of the likelihood of the disease in question being present[288]. Blinding was not possible in IP1-PROSTAGRAM as targeted biopsy was required of MRI lesions. Instead, the bias was reduced by requiring operators to follow a standardized biopsy procedure. Individual training was provided to all operators before commencing the

trial. In addition, the radiologists who reported the index test (MRI) were not present during the biopsy to ensure independence of the index and reference test.

Fourth, the diagnostic accuracy calculators may have been affected by the 'healthy volunteer' effect which can lead to a different risk profile in the study population to the community. To minimize spectrum bias, a screening test should be evaluated using the population in which it is planned to be used. To minimize this bias, men were not eligible to participate if they had a PSA test in the previous two years; further, a community-based recruitment strategy was conducted to ensure a representative sample of men from the local community.

7.6 Conclusion

This chapter has presented estimates of sensitivity and specificity using a number of well-established statistical methods to correct for verification bias. The point estimates of sensitivity at different histological thresholds suggests that a screen-positive fast MRI may have higher performance compared to PSA. The point estimates for specificity suggest that MRI Score ≥ 4 has sufficient levels to be considered for further evaluation as a screening test and did not differ from PSA.

It is clear that each MRI threshold has a trade-off between sensitivity and specificity. Given the findings of the previous chapter, an MRI score ≥ 4 is still recommended as the threshold to denote a screen-positive fast MRI unless strategies can be developed to manage the high prevalence of indeterminate MRI lesions in the general population.

Chapter 8 – Perceived patient burden and acceptability of fast MRI

8.1 Overview

The effectiveness of a screening tests depends on not only the sensitivity and specificity for significant prostate cancer but on acceptability in the general population. While the early chapters of this thesis have focused on clinical performance, this chapter will evaluate differences in patient experience and preference for PSA and fast MRI. Drawing on a similar methodology utilised to evaluate the acceptability of colorectal cancer screening tests, the IP1-PROSTAGRAM trial included a series of acceptability questionnaires designed to provide an unbiased and validated comparison of acceptability of PSA and fast MRI as screening tests.

8.2 Introduction

The acceptability of a screening test is an important determinant of participation in screening and fundamental to the effectiveness of a screening programme. If a screening test has low acceptability the attendance rate at initial and subsequent screening rounds can be poor[289]. The impact of attendance rates on the eventual effectiveness of PSA screening has been demonstrated in micro-simulation models within the ERSPC screening studies[290]. The relative reduction in prostate cancer mortality was almost eliminated when attendance to PSA screening was reduced to 30%.

A real-world example of this has been seen in colorectal cancer screening, where the most common early screening test was a guaiac based faecal occult blood (FOBT) which required individuals to apply multiple sample of faeces onto a test card with a spatula. Even though FOBT was effective at reducing mortality in randomised controlled trials[291], the uptake rates were sub-optimal when rolled out as a national screening test. The introduction of the Faecal Immunochemical Test (FIT) which was easier to complete and perceived as less unpleasant led to an increase in participation[292].

From a public health perspective, when multiple screening tests are available, the optimal screening choice is not always the most accurate test. Policy makers need to choose the test which offers the best possible balance between accuracy and uptake. There are a number of procedural differences between fast MRI and PSA which might impact acceptability and uptake of each test. PSA screening requires taking a blood sample which involves a short,

localized and sharp painful stimulus during needle insertion. On the other hand fast MRI requires lying still in an small, enclosed space for a longer period. There have been no previous studies comparing the acceptability for men undergoing PSA or MRI for prostate cancer. Studies in colorectal cancer screening have shown that acceptability can be related to expected and perceived burden of a screening test using domains such as the expected level of embarrassment or pain[293, 294].

Therefore, in this chapter I compare the expected and perceived burden of fast MRI and PSA as screening tests. A secondary aim was to quantify screening mode preference and examine predictors of overall burden in men having a screening MRI and PSA.

8.3 Methods

This study was embedded in the IP1-PROSTAGRAM trial[246] which has been described in detail in Chapters 5-6. Respondents to the recruitment strategy were informed that they would receive a series of tests to screen for prostate cancer. The initial invitation referred to a series of blood tests and imaging tests. It did not incorporate extensive details regarding each screening test to minimize selection bias from men who preferred to avoid a transrectal procedure.

8.3.1 Participants

Responding invitees received a standardised consultation with research staff to inform them about each screening test and to check exclusion criteria and/or contraindications for MRI or ultrasound. For the present study, we selected participants who completed acceptability questionnaires before and after PSA and MRI screening tests.

All responders received a detailed information leaflet to provide more information about each screening procedure and to facilitate informed decision making. The invitation letter have been described in Chapter 5 and the information leaflet is provided in Appendix III. Both were written and reviewed by GPs, urologists, radiologists and expert patient advisers. Further, this material was scrutinised by a Patient and Public Involvement panel and the UK NHS Health Research Authority and its associated Research Ethics Committee prior to gaining approval for this study.

8.3.2 Procedures

8.3.2.1 Fast MRI

The fast MRI examination was performed on either a 3T or 1.5T using a phased-array body coil but no endorectal coil. The median amount of time men spent in the scanner was 14 minutes 17 seconds (3T protocol) and 15 minutes 42 seconds (1.5T protocol). The full MRI protocol has been described in detail in Chapter 6.

Participants were asked to use the toilet to empty their rectum of stool and bowel gas prior to the MRI scan. An antispasmodic agent was administered via intramuscular injection to all participants to reduce motion artefact from bowel peristalsis. The standard drug was hyoscine butylbromine (Buscopan) 20mg. If this was contraindicated due to a history of glaucoma or cardiac arrhythmia, then Glucagon 1mg could be administered. This drug was administered prior to the patient entering the scan room and being positioned in the isocentre of the MRI scanner.

Participants were asked to change into a gown and remove any metallic objects prior to the scan. Weight and height were recorded prior to entering the scanning room. On entering the room participants were positioned supine on the scan table (Figure 38). A phased-array body coil was positioned over the pelvis. Noise defenders were provided with participants given a buzzer to contact the radiographer if required.

Figure 38: The MRI scan room for the 3T Siemen Magenetom Verio syngo MR B17 used in IP1-PROSTAGRAM.

During the MRI scan the radiographer recorded whether adequate images were taken. If this did not occur then the most relevant reason for this was recorded. This could either be due to poor tolerability where the participant was unable to tolerate the MRI scan and chose to discontinue despite the procedure being technically feasible (e.g. claustrophobia). Alternatively, due to technical factors where the procedure was not completed due to factors such as artefact from bowel gas. If there was overlap between these factors it was in the radiographer's discretion as to which was the most significant factor.

The radiographers collected additional information on the length of procedure defined as the time set up started (when participant enters the MRI room), time MRI commenced (time first localizer view was performed) and time MRI finished (time when the full MRI sequence was complete). Images deemed to be of insufficient quality were repeated and if the quality of the diffusion weighted imaging sequences was compromised by air, participants could be offered a rectal catheter to decompress the rectum.

8.3.2.2 PSA

All PSA samples were taken by an experienced phlebotomist in a designated phlebotomy room as shown in Figure 39. The approach for phlebotomy was based on the discretion of the phlebotomist taking into account the participant's preference from their previous experience. The most common approach used was Venepuncture using a butterfly needle. Participants were asked to remove clothing from the forearm to allow a tourniquet to be applied. This was applied to the forearm approximately 7-10 cm above the intended venepuncture site, moderately tight, with the radial pulse at the wrist still palpable. The tourniquet was kept in place no longer than one and a half minutes.

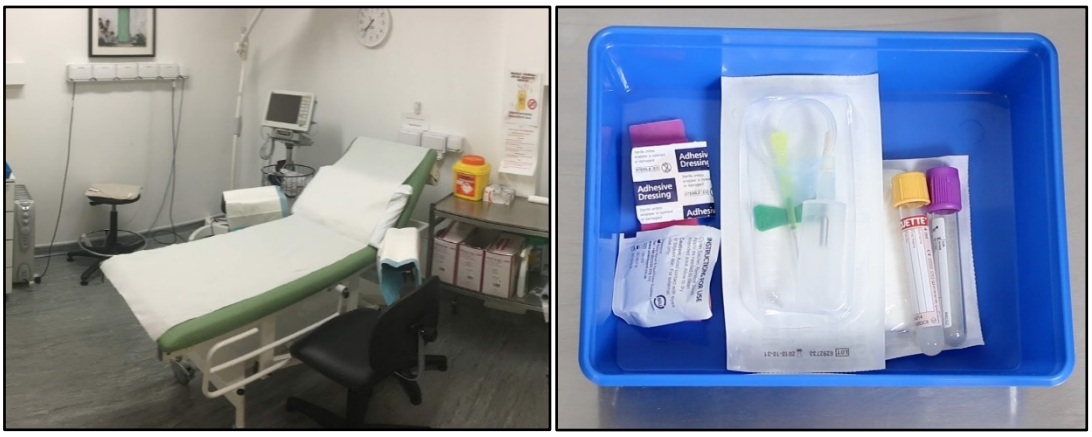


Figure 39: The phlebotomy room at one site from IP1-PROSTAGRAM and the equipment used for phlebotomy.

Two samples of blood were taken. A sample for PSA and a second EDTA sample for additional epigenetic testing in a pre-approved sub-study within IP1-PROSTAGRAM evaluating alternative serum-based biomarkers. After blood collection was complete a clean cotton dressing was applied over the skin at the insertion site. For each participant a standardised PSA sample worksheet was completed. This included detail on the reason if the procedure could not be completed which could either be related to tolerability or technical factors such as the procedure not being completed due to the factors such as inability to locate adequate veins. If there were overlapping factors it was at the operator's discretion which was the most significant factor.

Length of procedure was split into

- Time set up started: Defined as the time when set up of equipment commenced including patient preparation time
- Time phlebotomy commenced: Defined as the time phlebotomist applied tourniquet and started looking for veins
- Time phlebotomy finished: Defined as the time when the dressing was applied not including time to label, store and transport the samples.

8.3.3 Questionnaires

Participants completed two validated questionnaires before and after each screening test. The Expected Burden Questionnaire (EBQ) and Perceived Burden Questionnaire (PBQ) have been developed for use in bowel cancer screening. These have been used in studies investigating the acceptability of bowel cancer screening tests and were adapted for the screening tests in this study.

8.3.3.1 Expected burden questionnaire (EBQ)

Participants completed the EBQ questionnaire in the waiting area prior to having any screening test. The EBQ is comprised of four domains addressing the expected extent of embarrassment, pain, burden and anxiety caused by each test. This is followed up by a question summarising which test the patient expects to prefer. All items are on a 5-point Likert scale (1 = Not at all, 2 = slightly, 3 = somewhat, 4 = rather, 5 = extremely).

Each question, representing each component, was reported separately. An overall burden score was also calculated for each screening test by summing the response scores to the first four questions. Lower overall scores represent lower expected overall burden for that particular screening test.

8.3.3.2 Perceived burden questionnaire (PBQ)

*A*fter each screening test, participants completed a second questionnaire. This was completed in the waiting area immediately after having each test. At the point of completing the questionnaire participants did not know the final result of the screening procedure.

Similar to the EBQ, the PBQ comprises of five questions addressing the embarrassment, pain, burden and anxiety experienced from each test, and how likely the patient was to have the test again, if recommended. This was followed up by a question summarising which test the patient preferred. For each of the screening tests, 5-point Likert scales were used to elicit the subject's perception of anxiety, burden embarrassment and pain. The responses to each question was coded as 1 = not at all, 2 = slightly, 3 = somewhat, 4 = rather and 5 = extremely.

8.3.3.3 Supplementary questionnaires

In addition to the EBQ and PBQ, baseline questionnaires collected information on demographics such as educational levels and previous experience of prostate cancer screening. Each participant's subjective evaluation of prostate cancer risk was assessed using the cancer worry scale (CWS) which is a validated four item questionnaire to measure worry about the risk of developing cancer and the impact of worry on daily functioning. The short form Spielberger state-trait anxiety inventory (STAI-6) questionnaire was used to measure baseline anxiety in participants.

Clinical questionnaires were completed which included the Charlson Comorbidity Index (CCI) to assess the number of comorbidities per patient. This questionnaire has seven questions, each with various sub-questions. Weights are assigned for each condition. The total score equals the sum of the weights for the comorbidities the patient has. The final questionnaire was the IPSS (international Prostate Symptom Score) score which includes seven questions concerning urinary symptoms and one question concerning impact of urinary symptoms on quality of life. The responses to the questions concerning urinary symptoms range from 0 to 5, indicating increasing severity of the particular symptom.

8.3.4 Outcomes

The primary outcome was to compare the post-test burden (PBQ) for PSA and fast MRI. This includes the overall burden score calculated for each screening test by summing the response scores to each domain (anxiety, burden, embarrassment and pain). Individual comparisons of each domain were also conducted as part of the primary outcome.

Secondary outcomes included a similar comparison with pre-test burden (PBQ) and an examination of mean EBQ and EBQ scores across each domain. Additional outcomes were to compare the outcomes of each domain before and after each test, and the distribution of screening mode preference was reported. The final secondary outcome was to identify predictors of overall burden for each test.

8.3.5 Statistical Analysis

Statistical analysis was performed using with R version 4.0.3 (R foundation for Statistical Computing). Variables were initially inspected for being normally distributed. None of the variables were found to satisfy assumptions of normality required for paired t-tests. Multiple transformations, including log transformation, were implemented and tested for normal distribution. No transformation was found that was successful in transforming the distribution of variables into a normal distribution. Summary statistics of acceptability variables are presented using the median and IQR since the variable distributions are skewed.

Due to the paired nature of the data the primary analysis used a two-sample paired Sign test which is not limited by the distributional assumptions of the parametric paired samples t-test. This test has reduced power when compared with the Wilcoxon signed rank test as it is based on the direction of the sign of the observations, rather than their numerical magnitude. The Sign test looks specifically at the median value of differences and is not affected by the distribution of the data. The null hypothesis of the two-sample paired Sign test is that the median of the difference is zero.

Each question, representing each domain, is reported separately. An overall burden score was calculated for each screening test by summing the response scores to the first four questions, excluding the question relating to repeat test recommendation. Lower overall scores represent lower perceived overall burden for that particular screening test.

Secondary analysis used the Wilcoxon ranked sign tests which is a non-parametric test that does not assume normal distribution of the data. The Likert scores were assumed to be a continuous distribution to allow a comparison of the mean scores for anxiety, burden, pain and embarrassment.

Sankey charts were adopted to visually show the transitions of EBQ and PBQ scores between screening tests. This paired analysis included only men who completed both questionnaires. Spearman's rank correlation rho was performed to determine the correlation between pre and post scores. Differences in overall preference between tests were compared with the Chi Squared test.

A series of univariate regression analyses were conducted to evaluate determinants of overall burden of MRI. The overall burden was calculated by summing the Likert score of each component domain (anxiety, burden, embarrassment and pain). The score was dichotomised by the median value for each screening test so that outcome variable was binary.

Covariates were divided into three groups which include baseline factors such as age (continuous), Afro-Caribbean ethnicity, Index of Multiple Deprivation (IDDM), qualification at degree, A-level or equivalent, employment status, family history of prostate cancer and previous screening for prostate cancer. The IDDM is an area-based proxy for socioeconomic status and was dichotomised into the Least/Most Deprived. Psychological factors included pre-test levels of expected anxiety, burden, embarrassment and pain as measured by the EBQ. Procedural factors included variables which might affect the degree of burden for each test and included length of procedure, body mass index and prostate volume for MRI only.

Multivariate binary logistic regression was then performed for all variables in the univariate analysis with a significant or near significance $p < 0.1$. Statistical significance was set a $p < 0.05$.

8.4 Results

8.4.1 Study population

Between October 2018 and May 2019, a total of 411 men aged 50-69 years attended for screening. We excluded participants (n=3) for who had a history of prostate biopsy and a contraindication for fast MRI. Prior to screening, 403/408 (98.7%) of participants completed each domain of the pre-test EBQ.

All men completed the PSA screening test and the post-test questionnaire (PBQ). There were two men who did not tolerate the full MRI procedure who were retained in this questionnaire analysis. The mean duration of each screening test was 1.3 minutes for PSA (SD 0.59) and 19.7 minutes for MRI (SD 3.83). There was one participant who did not complete the post-test MRI questionnaire for logistical reasons (407/408, 99.7%).

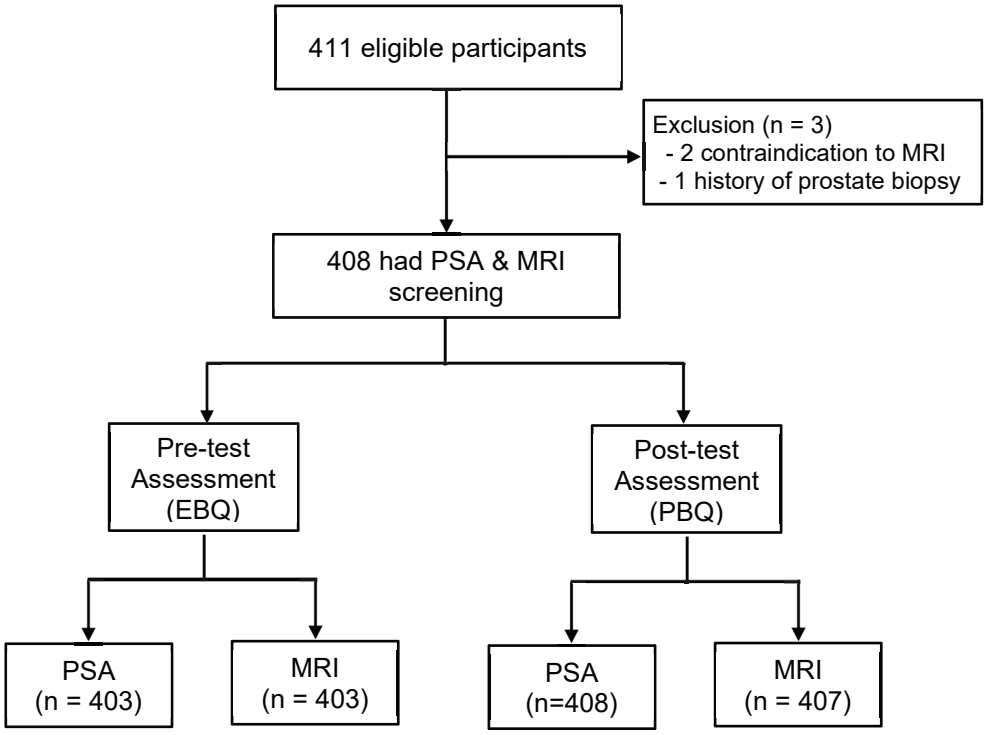


Figure 40: Overview of responses to the expected (EBQ) & perceived burden (PBQ) questionnaires among participants of received PSA and MRI screening in the IP1-PROSTAGRAM clinical trial.

Primary outcome

For the primary outcome, the perceived burden for MRI and PSA were compared. In total 30.8% perceived MRI to be worse than PSA, 18.7% perceived MRI to be better than PSA, and 50.4% perceived MRI and PSA to be the same (Table 38). The proportion of positive and negative differences from the Sign tests suggest that PSA was perceived to have a lower burden than MRI (p = 0.0007).

Table 38: Output of non-parametric two-sample paired Sign test comparing PBQ scores

	MRI +ve	PSA +ve	No difference	P-value (n)
Anxiety	108 (26.5%)	43 (10.6%)	256 (62.9%)	<0.0001*** (n=151)
Burden	75 (18.5%)	15 (3.7%)	316 (77.8%)	<0.0001*** (n=90)
Embarrassment	21 (5.2%)	7 (1.7%)	379 (93.1%)	0.013* (n=28)
Pain	9 (2.2%)	89 (21.9%)	309 (75.9%)	<0.0001*** (n=98)
Overall	125 (30.8%)	76 (18.7%)	205 (50.5%)	0.0007*** (n = 201)

*Significant at 0.05 level (two-sided) ***Significant at 0.001 level (two-sided)

In terms of the component scores, 108 (26.5%) men had an anxiety score for MRI larger than their score for PSA, 43 (10.6%) had a score for MRI less than their score for PSA, and 256 (62.9%) had a score for MRI equal to their score for PSA (p<0.0001). We can reject the null hypotheses that the medians of the differences in PBQ anxiety component scores between MRI and PSA are zero at a significance level of 0.0001. This implies that MRI had a higher anxiety component compared to PSA.

Similar to anxiety, for the burden component score, 75 (18.5%) perceived MRI to be more burdensome than PSA, 15 (3.7%) perceived PSA to be more burdensome than MRI, and 316 (77.8%) men perceived MRI and PSA to be the same in terms of burden. This difference in positive and negative proportion meant the null hypothesis could be rejected at a significant level of 0.0001. The results imply that MRI had a higher level of embarrassment compared to PSA. The embarrassment component showed that 21 (5.2%) of men perceived MRI to be more embarrassing than PSA, 7 (1.7%) perceived PSA to be more embarrassing than MRI, and 379 (93.1%) perceived MRI and PSA to be the same in terms of embarrassment.

The pain component was the only score which was higher for PSA. For pain, 9 (2.2%) perceived MRI to be more painful than PSA, 89 (21.9%) perceived PSA to be more painful than MRI, and 75.9% perceived MRI and PSA to be the same in terms of pain. Using the

proportions of positive and negative differences implies that PSA was perceived as more painful than MRI (p <0.0001). The outcomes of these are visualised in Figure 41.

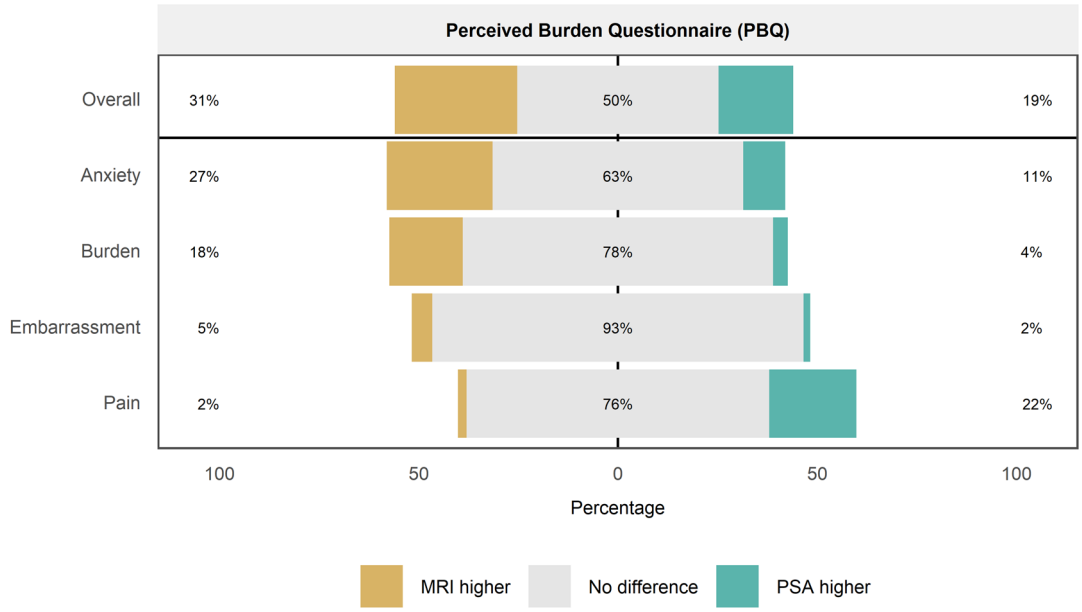


Figure 41: Divergent stacked bar charge showing the PBQ components for MRI vs PSA.

8.4.3 Secondary outcomes

8.4.3.1 Mean Scores and Wilcoxon Comparison

EBQ: The overall mean EBQ score for MRI and PSA was 1.45 (SD 0.65) and 1.46 (SD 0.54) respectively. This indicates that a larger proportion of men having PSA expected the procedure overall to be more burdensome than MR (1.80 vs. 1.85 p = 0.03). In terms of component scores of anxiety, burden, embarrassment and pain for MRI these were 1.75, 1.47, 1.32 and 1.26 respectively. For PSA, the component scores for anxiety, burden, embarrassment and pain were 1.63, 1.33, 1.21 and 1.67 respectively. Comparison between these scores showed that a higher proportion of men having MRI expected to have more anxiety (1.75 vs 1.63, p=0.02), to be more burdensome (1.47 vs 1.33 p<0.001) and have more embarrassment. The expected pain score for PSA was higher than MRI (1.67 vs. 1.26 p<0.001).

PBQ: The participants perceived the overall burden of MRI to be higher than PSA (1.21 vs. 1.16, p = 0.003). In terms of component scores of anxiety, burden, embarrassment and pain for MRI these were lower than the pre-scores at 1.46, 1.29, 1.08 and 1.04, respectively. For PSA the component scores were 1.26, 1.11, 1.03 and 1.25 respectively. A comparison of the component scores showed that the difference in overall score was due to a higher degree of

anxiety (1.46 vs. 1.26 p<0.01), burden (1.29 vs. 1.11, p < 0.001) and embarrassment (1.08 vs. 1.03 p=0.03). The pain score remained higher for PSA compared to MRI (1.25 vs. 1.04 p < 0.001).

Table 39: Wilcoxon signed-rank test to compare mean EBQ & PBQ scores

	MRI	PSA	P-value
Expected Burden (EBQ)			
Anxiety	1.75 (1.00)	1.63 (0.86)	0.02
Burden	1.47 (0.76)	1.33 (0.65)	<0.001
Embarrassment	1.32 (0.65)	1.21 (0.55)	<0.001
Pain	1.26 (0.67)	1.67 (0.72)	<0.001
Overall	1.45 (0.65)	1.46 (0.54)	0.027
Perceived Burden (PBQ)			
Anxiety	1.46 (0.77)	1.26 (0.65)	<0.001
Burden	1.29 (0.62)	1.11 (0.45)	<0.001
Embarrassment	1.08 (0.38)	1.03 (0.26)	0.03
Pain	1.04 (0.20)	1.25 (0.48)	<0.001
Overall	1.21 (0.35)	1.16 (0.36)	0.003

Figures are n (SD)

The mean EBQ and PBQ scores for PSA and MRI are detailed in Figure 42 which shows the trend towards higher anxiety, burden and embarrassment and lower pain scores with MRI. All component scores were lower in the PBQ compared to the EBQ suggesting that men expected both tests to be more burdensome than the reality.

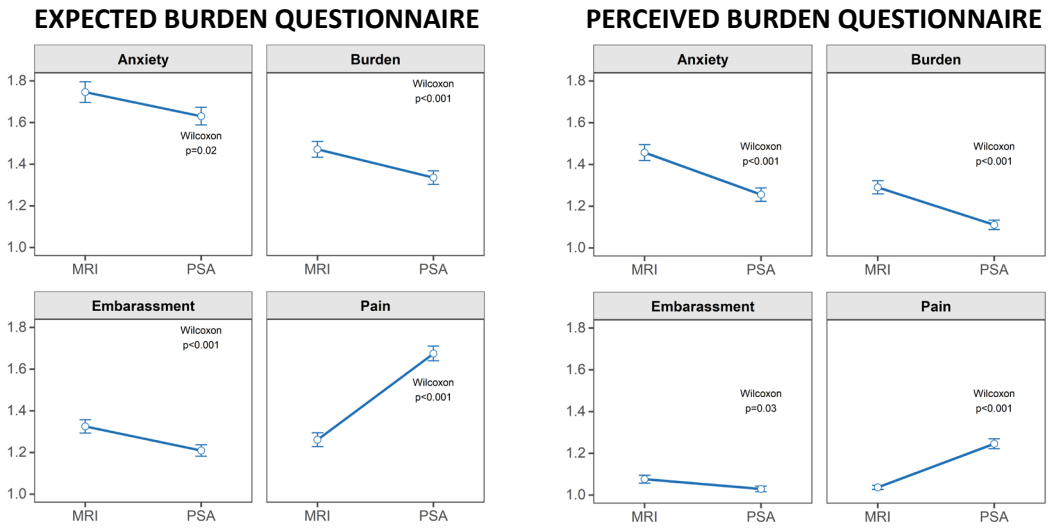


Figure 42: Line graphs with error bars shown the direction of mean EBQ and PBQ scores.

8.4.3.2 *EBQ scores*

The Sign test for the EBQ showed that 103 (25.6%) of men expected MRI to be worse than PSA, 143 (35.6%) expected MRI to be better than PSA, and 156 (38.8%) expected MRI and PSA to be the same. In contrast to EBQ the proportion of positive and negative differences from the Sign tests suggest that MRI was expected to have a lower overall burden than PSA ($p = 0.013$),

There was no difference in the expected difference in anxiety levels between PSA and MRI. 87 (21.6%) had an anxiety score for MRI larger than their score for PSA, 70 (17.4%) had a score for MRI less than their score for PSA, and 246 (61.0%) had a score for MRI equal to their score for PSA.

The expected pain component score was higher for PSA with 178 (44.2%) of men expecting PSA to be more painful than MRI compared to 28 (6.9%) expecting MRI to be more painful than PSA while 48.9% expected MRI and PSA to be the same in terms of pain

Table 40: Output of non-parametric two-sample paired Sign test comparing EBQ scores

	MRI +ve	PSA +ve	No difference	P-value (n)
Anxiety	87 (21.6%)	70 (17.4%)	246 (61.0%)	0.202 (n=157)
Burden	86 (21.4%)	51 (12.7%)	265 (65.9%)	0.0035** (n=137)
Embarrassment	76 (18.9%)	37 (9.2%)	290 (72.0%)	0.0003*** (n=113)
Pain	28 (6.9%)	178 (44.2%)	197 (48.9%)	<0.0001*** (n=206)
Overall	103 (25.6%)	143 (35.6%)	143 (35.6%)	0.013* (n = 246)

*Significant at 0.05 level (two-sided) ***Significant at 0.001 level (two-sided)

The components 'burdensome' and 'embarrassment' were higher for MRI. In total, 21.4% expected MRI to be more burdensome than PSA, 12.7% expected PSA to be more burdensome than MRI, and 65.9% expected MRI and PSA to be the same in terms of burden. This difference in positive and negative proportions meant the null hypothesis could be rejected at a significance level of $p=0.0003$. Similarly, for embarrassment, 76 (18.9%) of men expected the MRI to be more embarrassing than PSA, 37 (9.2%) of men expected PSA to be more embarrassing than MRI, and 290 (72.0%) of men expected MRI and PSA to be the same in terms of embarrassment.

8.4.3.3 MRI comparison of expected and perceived scores

The participant's transition between EBQ (pre-test) and PBQ (post-test) for MRI is shown in Figure 43 as a Sankey Diagram. There was a consistent trend for men to have lower pre-test scores across all component scores. In terms of anxiety, the reason for the increase was predominantly due to 17% of participants who had slight anxiety prior to MRI, then reporting no anxiety after completing the procedure. There were 16% men who expected MRI to cause some/rather/extreme anxiety prior to the procedure. After completing the MRI, this had reduced to 7%. Spearman's rank correlation between anxiety scores showed a moderate correlation value of 0.421 ($p<0.001$).

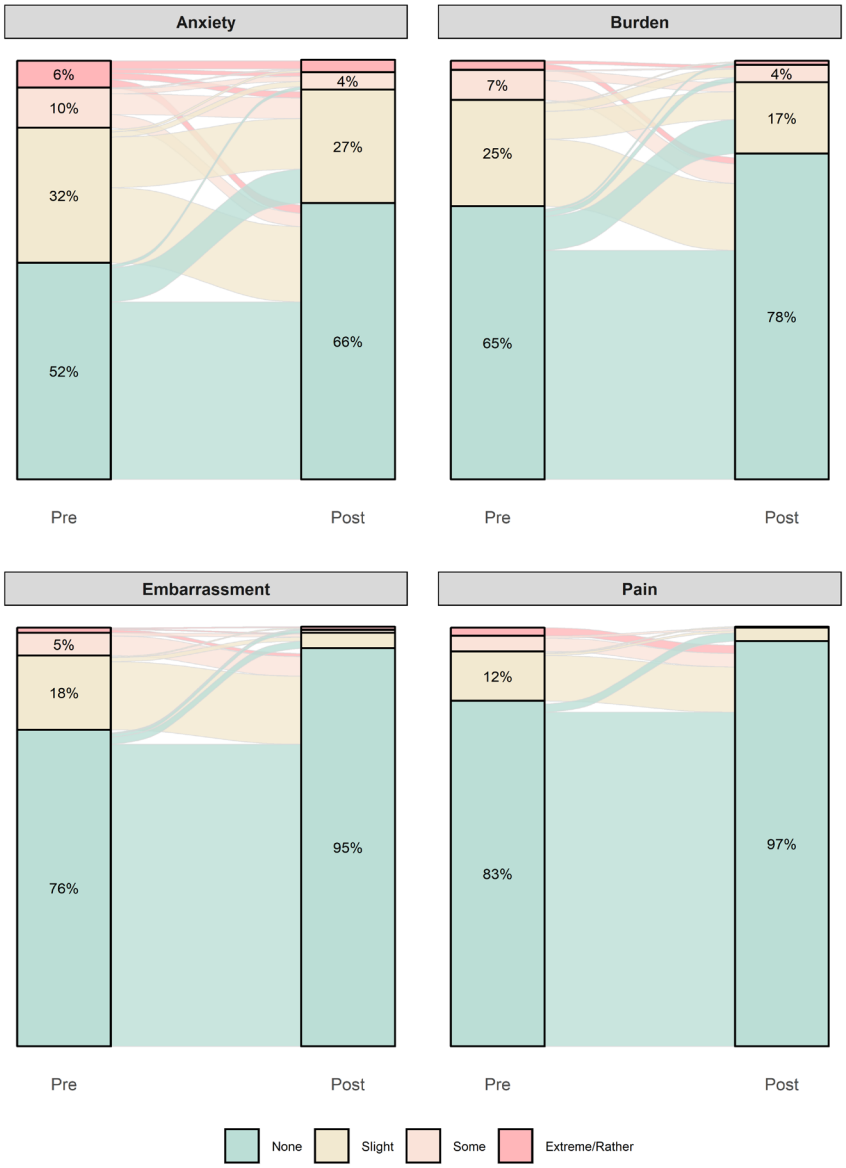


Figure 43: Sankey chart showing the relationship between the pre-test and post-test scores for MRI. Within each panel the left bar chart represents the pre (EBQ) score and the right represents the post-screening score. The ribbons connecting the left and right axis are proportional to the number of participants who transition from the pre-score to the post-score.

For the burden component, the proportion of participants who reported no burden increased from 65% to 78%. Similar to anxiety, this was predominantly caused by men who expected a slight amount of burden reporting that the MRI had less burden then expected. After completing the MRI, participants' reports of some/rather/extreme burden reduced to 5%. Spearman's rank correlation between anxiety scores showed a low correlation value of 0.227 ($p<0.001$).

For embarrassment, the MRI was perceived to be less embarrassing than expected. Prior to the scan, 24% expected it to have slight/some/rather/extreme embarrassment. After the procedure this was only 5%. A similar trend was seen for the pain component where 17% expected the MRI have slight/some/rather/extreme pain but only 3% reported experiencing any pain after completing the test. Spearman's rank correlation coefficients of 0.141 and 0.103 were observed for embarrassment and pain respectively ($p < 0.001$)

8.4.3.4 PSA comparison of expected and perceived scores

There was a similar trend for PSA with post-test (PBQ) scores improving compared to pre-test scores. For anxiety, the proportion of participants who reported no anxiety increased from 54% to 82%. Following the test, there were only 4% who reported some/rather/ extreme anxiety. The primary cause for the change was 26% participants who expected a slight amount of anxiety reporting no anxiety following the PSA test. Spearman's rank correlation between anxiety scores showed a low correlation value of 0.376 ($p<0.001$).

For the burden component, the proportion of participants reporting no burden increased from 74% to 92%, predominately due to a downgrading of slight pre-test scores. After completing the test, there were only 3% participants who reported the PSA test as causing some/rather/extreme burden. A similar trend was seen for embarrassment where the proportion of participants reporting no embarrassment was 98% after the PSA test. Spearman's rank correlation coefficients of 0.325 and 0.334 were observed for burden and pain respectively ($p < 0.001$).

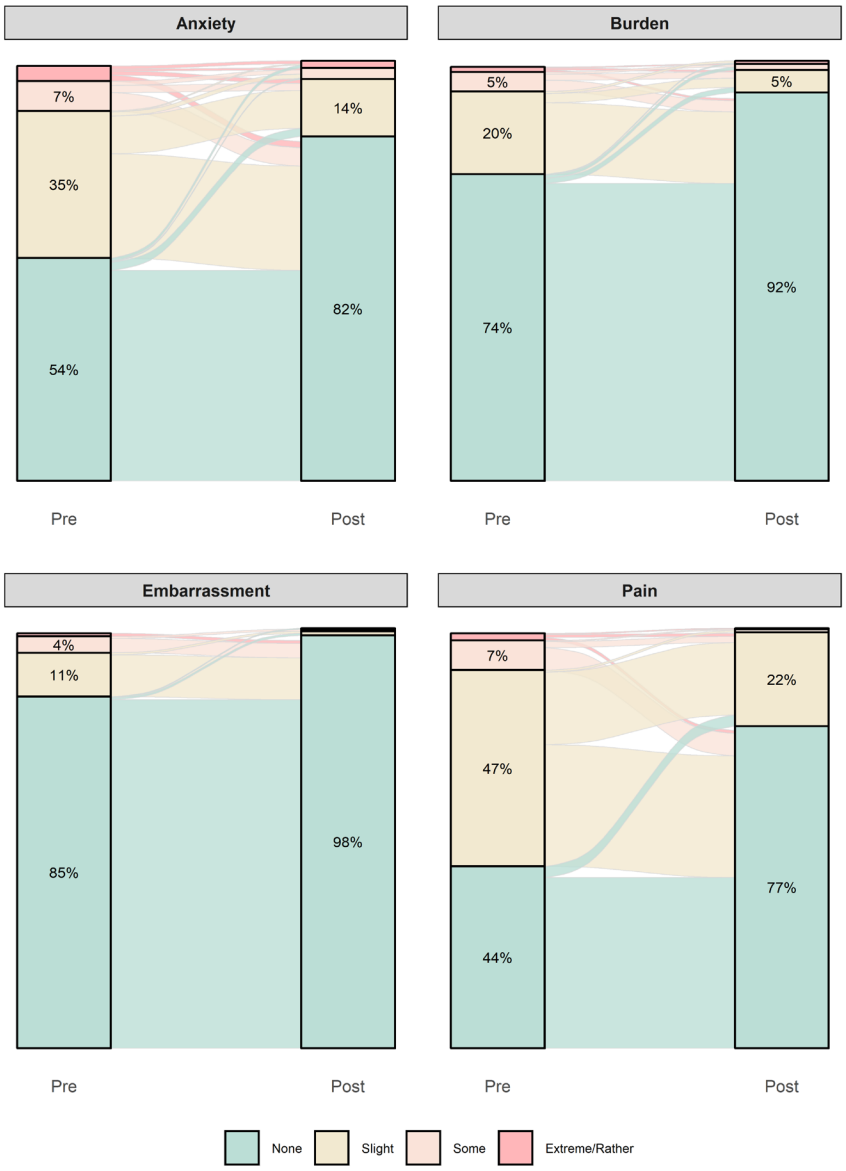


Figure 44: Sankey chart showing the relationship between pre-test and post-test scores for PSA.

The final component was the pain score which shows a similar pattern to the anxiety component. Prior to the test, 56% of participants who expected the phlebotomy procedure to be associated with slight/some/rather/extreme pain. After having the test this had reduced to 23%. There was only 1% remaining who reported the procedure as rather/extremely painful in contrast to 9% prior to having phlebotomy. Spearman's rank correlation between pain scores showed a low correlation value of 0.334 ($p<0.001$).

8.4.3.5 Preference for Screening Examination

Before the screening tests, the majority of participants had no preference on the type of screening examination. Of 408 participants, 194 (47.5%, 95% CI 42.7-52.4) had no preference, 106 (26.0%, 95% CI 21.9-30.5) preferred MRI and 79 (19.4% 95% CI 15.8-23.5) preferred PSA. This indicates that prior to screening, participants preferred MRI compared to PSA (+6.6%, 95% CI 8.4-12.3, p = 0.02). The proportion of participants who had no preference was higher than any category (p < 0.001). The difference between pre and post preference is shown in Figure 45.

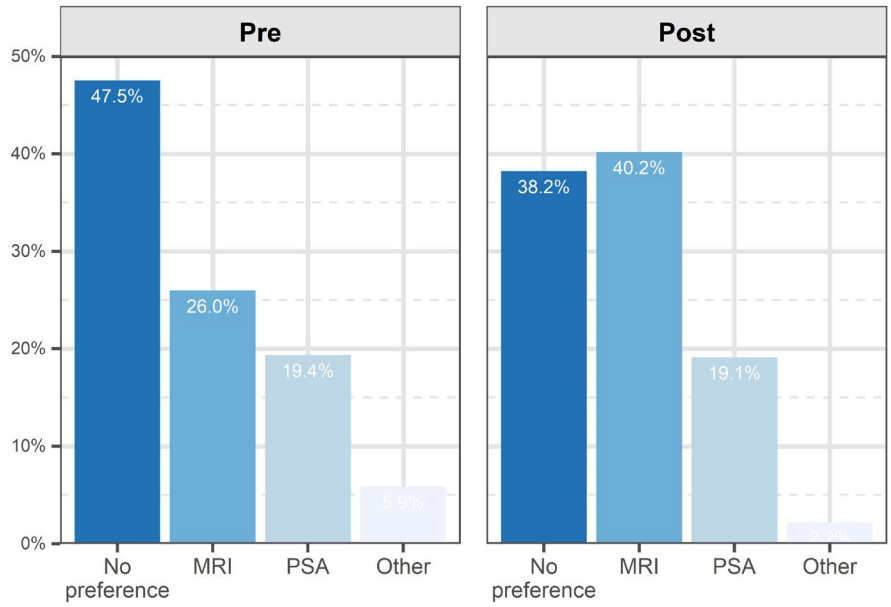


Figure 45: Bar charts showing participant's overall preference for screening examination.

After undergoing all screening tests, 164 (40.2% 95% CI 35.5-45.0) preferred MRI, 156 (38.2%, 95% CI 33.6-43.1) had no preference and 78 (19.1%, 95% CI 15.6-23.2) preferred PSA. The proportion of participants who preferred MRI compared to PSA was +21.1% (95% CI 14.9-27.1, p<0.001).

The shift in distribution of these scores is visualised in the Sankey flow chart (Figure 46). In men who expected to prefer MRI (27%), 16.4% continued to prefer MRI after the test, 7.1% switched to prefer PSA and 5.1% moved to no preference. Participants who had no preference prior to screening (48%) transferred to preferring either MRI (14.2%) or PSA (7.1%) after completing all tests. For participants who expected to prefer PSA (19%), 6.1% moved to MRI, 5.9% moved to no preference and 7.1% remained with PSA.

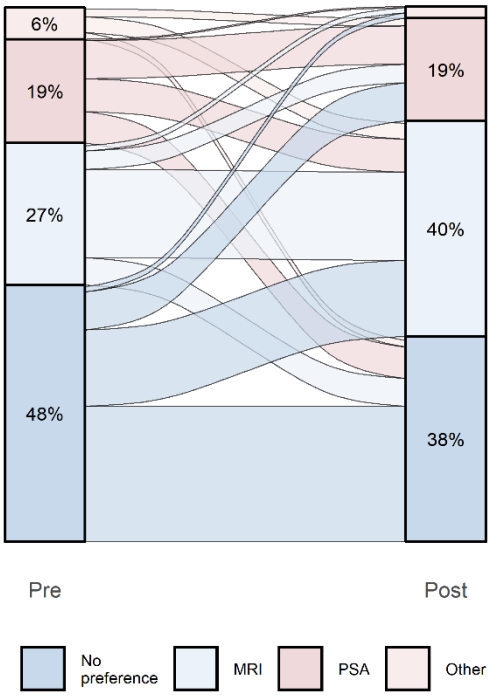


Figure 46: Sankey flow diagram focusing on the flow of overall preference before and after completing each screening test. Differences between figures in this chart and Figure 45 are due to rounding and the paired sample required for the Sankey Flow chart.

The correlation between overall preference and burden scores for PSA and MRI was assessed in Table 41. The correlation between preference for MRI and burden score was weakly or very weakly negative (-0.085 to -0.205). The correlation for PSA burden and test preference was also weakly negative (-0.005 - -0.07).

Table 41: Correlation between burden score & test preference

	MRI	PSA
Perceived Burden (PBQ)		
Anxiety	-0.181	-0.074
Burden	-0.156	-0.063
Embarrassment	-0.085	-0.054
Pain	-0.101	-0.005
Overall	-0.205	-0.067

Figures are Point-Biserial Correlation showing correlation between a dichotomous and a continuous variable

8.4.3.6 Determinants of overall burden

MRI: The outcomes of univariable analysis to examine the association between variables and presence of burden on MRI is shown in Table 42. Predictors that were not statistically significant at this stage were age, IDDM, qualification level, employment status, family history, MRI length, BMI and prostate volume. In the multivariable regression analysis, the presence of pre-test anxiety (odds ratio 2.59, p <0.001) and Afro-Caribbean ethnicity (odds ratio 0.521 p = 0.048) were identified as significant determinants of overall burden of MRI.

Table 42: Univariate and multivariate analysis for overall burden of MRI

	Univariate Analysis		Multivariate Analysis	
Covariate	**OR (95% CI)**	**p value**	**OR (95% CI)**	**p value**
Baseline Factors				
Age	1.01 (0.97 - 1.06)	0.611	-	-
Afro-Caribbean Ethnicity	0.62 (0.35 - 1.06)	0.086	0.521 (0.27 - 0.98)	**0.048**
Index of Deprivation*	1.16 (0.46 – 2.68)	0.731	-	-
Qualification Level**	1.24 (0.72 - 2.17)	0.443	- -	-
Employment Status¥	0.95 (0.54 - 1.71)	0.854	-	-
Family history^	1.81 (0.85 - 3.77)	0.434	-	-
Previous screening^^	1.20 (0.73 - 1.98)	0.243	-	-
Psychological Factors				
Expected Anxiety	2.59 (1.95 - 3.55)	<0.001	2.16 (1.44 – 3.33)	**<0.001**
Expected Burden	2.88 (2.03 - 4.23)	<0.001	0.979 (0.784 – 2.42)	0.26
Expected Embarrassment	2.21 (1.54 - 3.25)	<0.001	0.485 (0.485 - 1.53)	0.64
Expected Pain	2.58 (1.75 - 4.00)	<0.001	0.725 (0.725 - 2.25)	0.41
Procedural Factors				
MRI Procedure Length	1.06 (0.99 - 1.13)	0.105	-	-
BMI	1.01 (0.94 - 1.07)	0.834	-	-
Prostate Volume	1.00 (0.97 - 1.02)	0.766	-	-

- Insufficient association on univariate analysis to continue in the multivariate analysis
* Most Deprived (Quintile 5) vs. Other, ** A-Level, Degree or Equivalent vs. lower/none
¥ Employed vs. any other. ^ First or second degree relative with prostate cancer,
^^ Defined as either PSA or DRE screening

PSA: For the univariable analysis, statistically significant variables were Afro-Caribbean ethnicity (odds ratio 1.8, p-value 0.05), qualification level (odds ratio 0.54, p-value = 0.04), pre-test anxiety, pre-test burden, pre-test embarrassment and pre-test pain (Table 43). Following multivariate analysis, the presence of pre-test anxiety (odds ratio 1.12, p <0.001) and pre-test expected burden (odds ratio 1.14 p < 0.001) were identified as significant determinants of overall burden.

Table 43: Univariate and Multivariate analysis for overall burden of PSA

Covariate	Univariate Analysis		Multivariate Analysis	
	OR (95% CI)	**p value**	**OR (95% CI)**	**p value**
Baseline Factors				
Age	0.99 (0.93 - 1.04)	0.667	-	-
Afro-Caribbean Ethnicity	1.80 (1.98 - 3.28)	0.054	1.03 (0.95-1.12)	0.47
Index of Deprivation*	0.53 (0.12 – 1.58)	0.318	-	-
Qualification Level**	0.54 (0.30 – 0.98)	0.041	0.932 (0.86-1.01)	0.075
Employment Status¥	1.50 (0.74 - 3.31)	0.284	-	-
Family history^	0.69 (0.23 - 1.70)	0.468	-	-
Previous screening^^	0.96 (0.53 - 1.73)	0.901	-	-
Psychological Factors				
Expected Anxiety	2.84 (2.05 - 4.09)	<0.001	1.12 (1.06 - 1.18)	**<0.001**
Expected Burden	3.55 (2.34 - 5.55)	<0.001	1.14 (1.06 - 1.22)	**<0.001**
Expected Embarrassment	1.97 (1.28 - 3.03)	0.002	0.959 (0.89-1.03)	0.19
Expected Pain	3.26 (2.13 - 5.19)	<0.001	1.05 (0.98 - 1.12)	0.17
Procedural Factors				
PSA Procedure Length	0.95 (0.56 - 1.53)	0.843	-	-
BMI	0.96 (0.89 - 1.03)	0.295	-	-

- Insufficient association on univariate analysis to continue in the multivariate analysis
* Most Deprived (Quintile 5) vs. Other, ** A-Level, Degree or Equivalent vs. lower/none
¥ Employed vs. any other. ^ First or second degree relative with prostate cancer,
^^ Defined as either PSA or DRE screening

8.5 Discussion

8.5.1 Principle Findings

This chapter has focused on patient reported measures related to each screening test with more than 6,483 responses across the EBQ and PBQ domains. We have shown that the overall level of burden for fast MRI and PSA was minimal. Few men reported high levels of anxiety, burden, embarrassment or pain following either test. Participants indicated an overall preference for fast MRI after completing all screening tests despite comparison of the individual domains showing that anxiety, burden and embarrassment were higher for fast MRI.

Key strengths of this study are the high response rate, well-balanced cohort, paired design and use of validated patient-reported outcomes measures. The paired design allowed a direct comparison between the tests and the response rate to both questionnaires was above 98%. The generalisability of the results was enhanced by the broad representation of ethnic groups, socio-economic status and educational backgrounds within IP1-PROSTAGRAM in contrast to previous studies which have recruited predominantly Caucasian males from a narrow educational and social background. This was due to a purposive sampling strategy which helped promote diversity within the trial's recruitment strategy.

The domain level findings of higher rates of post-anxiety, burden and embarrassment associated with fast MRI could be attributed to certain features of the MRI procedure. The experience of having an MRI was less familiar to the majority of participants and likely associated with higher anxiety levels. An MRI had a longer duration and required participants to change into safe hospital-provided MRI compatible clothing so is likely to be scored as higher burden and embarrassment. In contrast, PSA is a rapid test requiring minimal exposure; and the process of having a blood test will be familiar to men within this age category.

Although the differences between fast MRI and PSA were statistically significant, the actual differences were often small when considered relative to the score within each domain. For example the overall perceived burden score was 1.16 for PSA and 1.21 for MRI. This difference of 0.05 on a Likert scale was statistically significant ($p = 0.03$) but unlikely to be of practical clinical significance. Previous studies have suggested that the threshold for clinically important differences in patient reported outcome measures is approximately 0.5 SD[295]. In this chapter the pooled SD was 0.35 for MRI and 0.36 for PSA for overall scores which

suggests that the variation in scores is unlikely to have a clinically significant impact for the majority of participants. This supports the conclusion that acceptability was high for both tests.

It should be noted that participants reported a clear preference for fast MRI after completing all screening tests. Out of 408 participants, 164 preferred fast MRI, 156 had no preference and only 78 preferred PSA. Given that the overall burden score was calculated as an average across each score and not weighted towards a particular score, it is possible that participants value pain caused by a screening test to be a more significant factor that other domains in the burden questionnaires.

Another interesting finding, given the prevalence of prostate cancer among Afro-Caribbean men was the racial differences identified in the evaluation of predictors of screening test burden. For fast MRI, Afro-Caribbean ethnicity was associated with a lower burden on univariate (OR 0.62) and multivariate analysis (OR 0.52). The significance of this finding on multivariate analysis was borderline ($p = 0.048$) so should be treated with caution.

Pre-test MRI anxiety explained most variance in the MRI multivariate models. Other studies have identified pre-test anxiety as important predictors of patient experience[296]. An important implication is that the patient experience may be primarily determined by the initial perception of fast MRI rather than background or procedural factors. The implications of this finding are discussed in the following section.

Several changes regarding the pre and post domain scores for anxiety, burden, embarrassment and pain were noticed. First, it was observed that both PSA and MRI were assigned higher burden scores by participants prior to screening. After completing the test, the majority of participants rated the tests as the same or lower burden than expected. This means that, in contrast to participants expectations, MRI and PSA caused less anxiety, burden, embarrassment and pain than expected. Second, despite extensive pre-test counselling, there was a proportion of men who reported that MRI and PSA was expected to be embarrassing. This was not reflected in reality as following both tests, minimal men reported finding the tests embarrassing which might suggest that men were not fully aware of the intricacies of each procedure prior to test.

8.5.2 Comparison to previous studies

There is a paucity of published data comparing patient reported outcomes for PSA and MRI as screening tests. There have been a few studies focused on anxiety levels in men with an abnormal PSA test which have generally found that receipt of an abnormal test did not have a significant effect on anxiety levels[297, 298].

Our findings can be compared with reports from other screening studies into imaging test for bowel cancer screening. Wijkerslooth et al [299] compared the PBQ scores of CT coloscopy and standard colonoscopy in a randomised controlled trial. The PBQ scores for these modalities are shown in Table 44 in comparison to the findings within the results section. The comparison suggests that MRI and PSA have lower levels of embarrassment, burden and pain to existing screening tests which have been evaluated using the same validated questionnaires.

Table 44: Comparison of PBQ with Wijkerslooth et al[299]

PBQ Score	Embarrassment	Burden	Pain
MRI^	1.1 (0.38)	1.3 (0.6)	1.0 (0.2)
PSA^	1.0 (0.26)	1.1 (0.5)	1.3 (0.5)
CT Colonoscopy*	1.5 (0.7)	2.0 (0.9)	2.1 (1.0)
Colonoscopy*	1.4 (0.6)	1.8 (0.9)	1.8 (1.1)

Figures are n (SD).
^ Rounded to 1 decimal for comparison with Wijkerslooth et al [299]
*Results from Wijkerslooth et al [299]

Wijkerslooth et al [299] noted a similar finding that although CT colonography had a higher level of burden across multiple domains, there was no impact on the rate that participants would recommend the test as an overall preference level. A similar finding was noted in this chapter that men reported MRI as the preferred test despite high burden score compared to PSA.

8.5.3 Implications of findings

This analysis of patient reported experience measured (PREMS) for fast MRI will be valuable in further understanding the role of MRI as a new screening test. I have identified a number of potential psychological barriers to fast MRI and PSA screening. The multivariate analysis found that expected anxiety and/or burden are important patient-related determinants of screening test experience. The fact that pre-test perceptions of the test strongly influenced the experience indicates that more intensive effects may be warranted to address preconceptions and patient-related anxiety for each test.

The participants in IP1-PROSTAGRAM received extensive explanation and discussions as part of the consent process and it is less likely that improvements can be made by providing more verbal or written or visual (e.g., video) information prior to fast MRI. Other new screening tests, such as low dose-CT for lung cancer, have implemented alternative strategies to support having a CT scan[300]. Studies have shown that video or online interventions can be effective at improving anxiety and knowledge prior to imaging tests [301]. Video intervention improved knowledge transfer to a greater degree than print materials particularly in populations with lower levels of health literacy. The advantage of video as a format for delivering the information is the scalability and delivery of the technique for use in a screening setting. Further work is needed to develop and evaluate similar strategies for MRI screening.

That there were small but significant increases in embarrassment, burden and anxiety in men undergoing a fast MRI compared to PSA highlights the need for improvements to the fast MRI process to minimise the amount of expected discomfort and embarrassment. This is important given that patient experience of a screening test is an important determination of future and ongoing participation with multiple rounds of screening[302].

There are a number of areas which could be considered to reduce the overall burden of MRI. The fast MRI protocol in IP1-PROSTAGRAM was set up to be around 15 minutes but since the development of the protocol in Chapter 4, there are several faster techniques which have been developed to reduce scanning time without impacting diagnostic accuracy[303]. Fast MRI protocols with acquisition times from five to nine minutes have been shown in single centre studies to deliver a high sensitivity for significant prostate cancer[144, 304]. A fast MRI protocol with rapid scanning times could be more acceptable to men than a PSA pathway which often includes a digital rectal examination (DRE).

In addition, the use of a facility-provided gown for the MRI procedure was required to reduce metallic artifacts and risk of thermal burns from metallic components in clothing. Designing more acceptable MRI safe compatible garments would be a consideration to minimise embarrassment compared to the standard hospital gown. Another potential improvement to fast MRI might be in avoidance of intra-muscular buscopan and changing to oral agents which could be as effective at reducing bowel peristalsis if taken at an early stage before the MRI.

8.5.4 Limitations

A number of limitations should be acknowledged. First, an important limitation is that the EBQ and PBQ questionnaires have been validated for colorectal cancer screening so were adapted by this study for prostate cancer screening tests. There is a lack of validated questionnaires designed for prostate cancer screening tests so these were chosen as an alternative. However, we cannot exclude that important domains related to these prostate cancer tests were not captured by these questions. To account for this, we included additional generic questions related to overall preference.

Second, when determining overall preference for screening tests, participants were given no information on the predicted accuracy of fast MRI and PSA, although the participation information leaflet informed participants that the diagnostic accuracy of PSA and fast MRI in a screening setting was uncertain. Diagnostic accuracy is an important factor in patient preference of screening test choice[305]. As MRI is a high-tech novel imaging test, it is possible that the novel technology could have biased participants in favour of this modality. Indeed, the attraction of participating in the study may have been the fact that MRI was being evaluated leading to participants who were already keen to have the fast MRI scan.

Third, although participants received identical information leaflets and a standardised consultation prior to each screening test, it is still possible that not all participants had a similar understanding of the fast MRI and PSA screening tests. As has been previously discussed, future work is needed to target improvements in information delivery particularly to increase awareness of the potential psychological burdens of embarrassment and anxiety associated with MRI.

Fourth, the overall preference responses were recorded immediately before and after each screening test as has been done in other studies. However, subsequent studies have shown that preferences may change over time[306]. After five weeks, participants may have less pronounced preferences so measuring experience after a longer period of time may be more predictive of future behaviour and attendance at screening. This design was considered but due to the multiple tests in IP1-PROSTAGRAM it was deemed that participants might find it difficult to differentiate tests as has occurred in other studies[307]. Future studies with fewer tests might benefit from longer follow-up questionnaire studies to confirm which factors are likely to constitute the basis of future decisions regarding preference for repeat screening tests.

Fifth, the non-random order of the tests may have introduced bias in the patient's answers. The protocol stipulated a fixed order to ensure that the PSA test was performed before the ultrasound examination which can induce false elevations of PSA levels. In addition, the PSA sample included an additional sample for epigenetic testing which could have increased the length of testing. However, as the majority of samples were taken by vacutainer the impact of an extra serum sample is likely to be minimal on timings.

8.6 Conclusion

The findings of this chapter show that fast MRI and PSA are both acceptable as screening tests for men aged 50-69 years. Both tests were associated with minimal amounts of anxiety, burden, embarrassment and pain. There was a small but significant lower overall burden for PSA testing. Interestingly, this did not lead to participants reporting an overall preference for PSA testing. Instead, the majority of participants preferred fast MRI which could indicate that patient compliance for this modality would be improved with this as a screening test.

Chapter 9 – Evaluation of screening pathways using a combination of MRI and PSA

9.1 Overview

In previous chapters the performance of fast MRI and PSA have been compared as independent tests. Although MRI as the primary screening test is attractive due to its high sensitivity for clinically significant disease, a combination with a more cost-effective serum-based biomarker could be an alternative approach to deliver the pathway on a large-scale. As a result, in the subsequent chapters of my thesis, I evaluate methods of combining PSA and fast MRI. This chapter examines different combinations of PSA and fast MRI to find a new screening pathway that provides a suitable balance of benefits and harms from screening.

9.2 Introduction

A general principle of screening is that the first test in a pathway should be the 'discovery test' with a high sensitivity for identifying the majority of people with the disease[308]. This discovery test can be set with a high sensitivity and corresponding low specificity provided there is a subsequent acceptable follow-up test to address false positives. Previous chapters have evaluated whether alternative tests could be used with MRI to reduce the false positive rate.

PSA represents a cheap and simple test and although it has a low sensitivity at 3ng/ml, previous studies have shown that the sensitivity of PSA could be improved by lowering the threshold[309]. There are many different combinations of PSA and fast MRI which could be followed and we compared 13 different pathways including PSA-alone, fast MRI-alone and a range of PSA thresholds and MRI scores. The performance of each pathway is evaluated focusing on trade-offs between biopsy referral rates, false positives and detection of clinically significant prostate cancer. The optimal threshold for PSA is considered in combination with MRI.

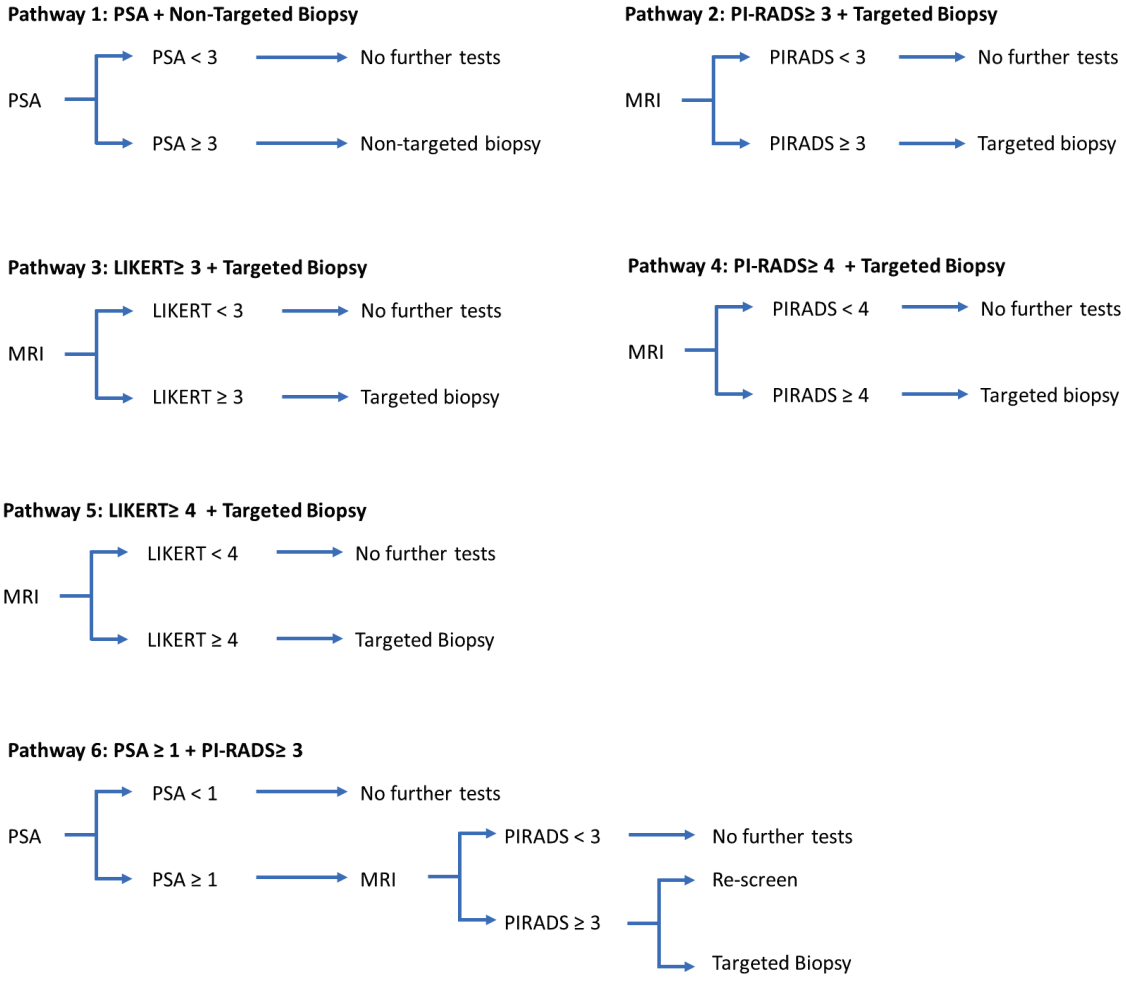
9.3 Methods

9.3.1 Study population

This study included data from the population-based IP1-PROSTAGRAM screening study. The design and primary outcomes of this study have been described in previous chapters. The details of each test have also been previously described in Chapters 6 and 8. If any test was screen-positive, men were advised to have prostate biopsy. In this way, we were able to therefore construct decision trees using combinations of the tests at various thresholds.

9.3.2 Screening pathways

Thirteen theoretical screening pathways were evaluated including a single test, or PSA and fast MRI in combination. The pathways can be considered in three groups starting from a single test screening strategy (Pathways 1 to 5) to a sequential pathway with PSA ≥ 1ng/ml as a first-line triage test (Pathways 6-9) and finally a group with PSA ≥ 3ng/ml as the first line test (Pathways 9-13). All pathways are summarised in Figure 50.



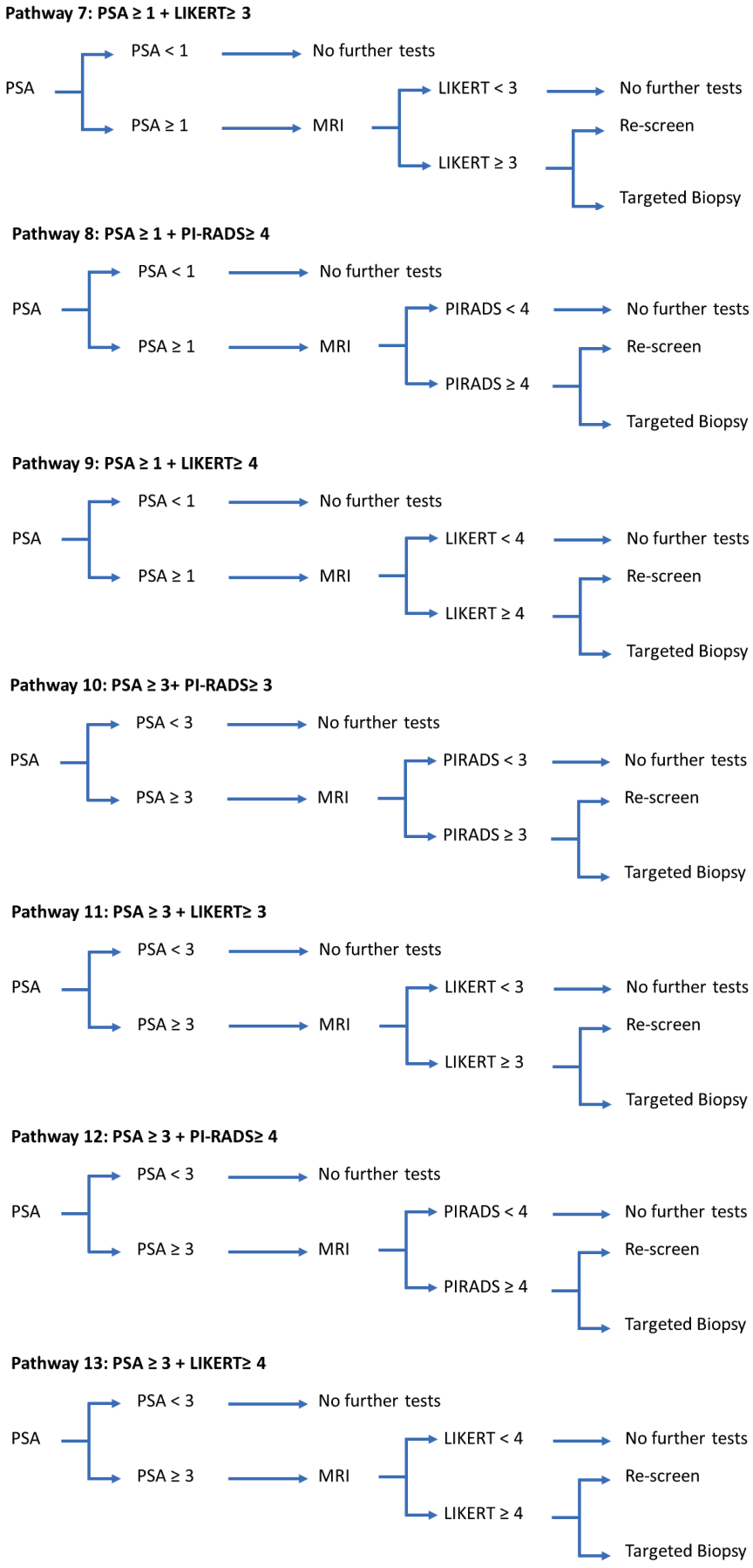


Figure 47: Schematic representation of 13 pathways for PSA and fast MRI as screening tests
Group 1 (Pathways 1-5) assumes that a single screening test either as PSA or MRI. Pathway 1 is considered the reference standard for the purposes of analysis. Group 2 (Pathways 6-9) used PSA ≥ 1ng/ml as an initial triage test prior to MRI and Group 3 (Pathways 10-13) increased the PSA threshold to ≥3ng/ml.

For the purpose of this analysis, Pathway 1 (PSA ≥ 3ng/mg) was designated the reference pathway as it reflects the current screening standard that has been evaluated in randomised controlled trials. In this pathway it was assumed that men underwent biopsy without image guidance to reflect the non-targeted nature of the pathway. All other pathways included the targeted biopsy cores only. A PSA threshold of PSA ≥ 1ng/ml (Pathways 6-9) was chosen on the basis of previous literature which suggests that this PSA threshold has high sensitivity for significant disease[86] and confers low risk of long term prostate cancer specific mortality[310].

9.3.3 Outcomes

The primary endpoints were a comparison of the biopsy and detection rates between each pathway and evaluation of the optimal threshold for PSA in each pathway. The biopsy rate was defined as the proportion of participants recommended for a biopsy due to an positive result from each pathway. The biopsy outcomes were dichotomised so that a true positive result was considered in men with any length of Gleason ≥ 3+4. The detection rate was defined as the proportion of participants found to have Gleason ≥ 3+4 during a biopsy directed by each pathway.

Secondary outcomes included a range of performance metrics such as overdiagnosis rate, intervention rate, positive predictive value (PPV), false positive rate, number needed to screen and number needed to biopsy. Overdiagnosis rate was defined as the proportion of participants with Gleason 3+3 during the biopsy. The compliance rate was the proportion of participants who received a recommendation for biopsy and underwent the procedure. The intervention rate was defined as the proportion of participants who underwent active and curative treatment for prostate cancer. Active surveillance was not included within this definition.

The PPV was defined as the proportion of participants with a positive pathway who subsequently underwent biopsy and was found to have Gleason ≥ 3+4. The false positive rate was defined as positive screening pathway with a biopsy which did not show Gleason ≥ 3+4. The number needed to biopsy was defined as the number of biopsies needed to find one participant with Gleason ≥ 3+4. The number needed to screen is the number of screening tests needed to find one participant with Gleason ≥ 3+4.

9.3.4 Statistical analysis

The analysis of biopsy rate in relation to significant cancer detection was performed for each of the thirteen pathways. Each endpoint was compared to the reference pathway (Pathway 1). The differences in proportions was calculated using the chi squared test or Fisher's exact probability test. The Wilson score method was used to obtain 95% confidence intervals.

The relationship between biopsy rate and cancer detection rate for each pathway was explored by fitting a locally weighted regression smoothing scatterplot (LOESS) line to cancer detection rate and biopsy rate. The reference pathway (Pathway 1) was excluded from the LOESS curve. LOESS regression is a non-parametric technique for fitting a smooth line to characterise the relationship between two continuous variables and indicate non-linear trends.

Receiver operating characteristic curve (ROC) analysis determined thresholds for PSA in pathways when combined with MRI. The threshold values were calculated using the Youden Index (maximum sensitivity and specificity with no constraint on sensitivity). Due to the need for high sensitivity of a screening pathway, PSA thresholds were also calculated for sensitivity constrained at 90% and 100%. Confidence intervals were calculated with the non-parametric bootstrap method using 1,000 bootstrapped datasets.

Descriptive statistics were used to show the baseline demographic information for participants. Means and standard deviations are reported for continuous variables with a normal distribution. For variables not normally distributed, medians and interquartile ranges (IQRs) are reported. All statistical analyses were performed with R version 2.3.1 (R foundation for Statistical Computing).

9.4 Results

The study population included 408 participants who were consented and eligible to receive a fast MRI and PSA test. The characteristics of the participants have been described in Chapters 6 (see Table 21).

9.4.1 Primary outcomes

Biopsy and Cancer detection rates: The LOESS plot showed a non-linear relationship between biopsy rate and cancer detection rate across the pathways (Figure 51). The curve illustrates the trade-off between biopsy rate and cancer detection rate across each pathway. There was a decreasing trend in the proportion of participants who underwent a biopsy observed from Pathways 2 to 13.

In comparison to the reference pathway (Pathway 1), there was a decrease in biopsy rate for Pathway 8 at -2.9 (95% CI -0.07-3.9, p=0.044), Pathway 9 at –4.7% (95% CI -8.6--0.8, p= 0.018), Pathway 10 at –7.6% (95% CI -11.1--4.1, p<0.001), Pathway 11 at –7.1% (95% CI -10.7--3.5, p<0.001), Pathway 12 at –7.6% (95% CI -11.1--4.1, p<0.001) and Pathway 13 at –8.1% (95% CI -11.5--4.6, p<0.001). There was net increase in biopsy rate of +7.6% (95% CI 2.6-12.6, p= 0.002) for Pathway 2 and +13.0% (95% CI 7.7-18.3, p<0.001) for Pathway 3. There was no statistical difference in the biopsy rate for Pathway 4, Pathway 5, Pathway 6 or Pathway 7.

There was a net increase in cancer detection rate for Pathway 2 at +2.2% (95% CI 0.0-4.4, p= 0.049) and Pathway 3 at +2.5% (95% CI 0.2-4.7, p=0.032). There was no significant difference in cancer detection rates in comparison to the reference pathway for the remaining pathways.

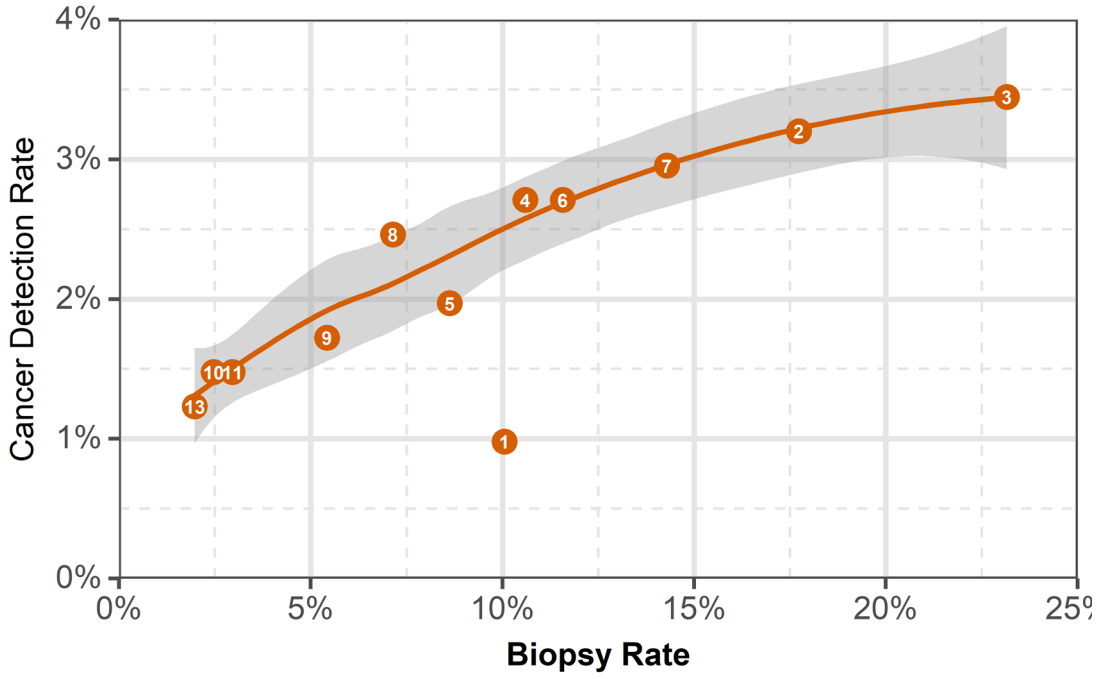


Figure 48: Local regression curve illustrating the non-linear trend (with 95% confidence intervals) between biopsy rates and significant cancer detection rates for each pathway. Each circle represents a pathway with cancer detection rate plotted against the biopsy rate. The reference pathway (Pathway 1) is excluded from the LOESS curve. Each pathway is labelled within the point. The shaded area represents the 95% confidence interval.

Optimal PSA Threshold: The pathways which include PSA as a triage test (Pathways 8-13) were the ones which reduced biopsy rates in comparison to the reference pathway. The cancer detection was increased across each pathway but it varied depending on the PSA threshold. Figure 52 shows the optimal cut-off for PSA for a pathway which includes PI-RADS ≥ 4. Depending on the criteria for the optimal threshold, this was 1.7ng/ml for maximising the Youden index, 1.02ng/ml when sensitivity was maintained at 90% and 0.9ng/ml to maximise sensitivity. The biopsies rates, sensitivity and specificity for each threshold are shown in Table 45.

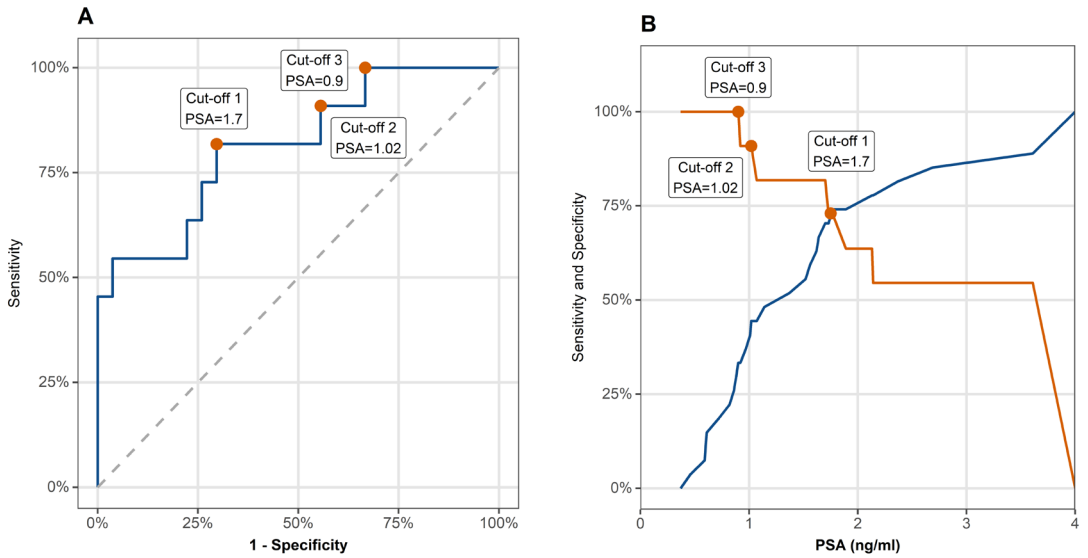


Figure 49: (A) Receiver Operating Curves for PSA based on 38 prostate biopsies conducted in participants with PI-RADS ≥ 4. It should be highlighted that these curves are made from men referred for biopsy on the basis of PI-RADS ≥ 4 and thus should be interpreted for this group. Different cut-offs are applicable for other pathways where biopsies were performed in different groups of men. (B) Sensitivity and specificity plot showing the optimal PSA thresholds and corresponding sensitivities and specificities.

Table 45: Optimal PSA cut-off for Pathways 8 and 12 (PI-RADS≥4)

Method	PSA optimal cut-off	Biopsy Rate	Sensitivity	Specificity
Youden	1.7ng/ml	4.7%	81.8%	70.3%
90% Sensitivity	1.02ng/ml	6.9%	90.9%	44.4%
Max Sensitivity	0.9ng/ml	7.8%	100%	33.3%

Youden = Sensitivity + Specificity - 1

The optimal PSA threshold varied depending on MRI score. For Pathways 6 and 10 (PI-RADS ≥ 3), using the maximum Youden method the optimal PSA threshold was 1.7ng/ml, to maintain sensitivity at 90% it was 0.9ng/ml and to maximise sensitivity it was 0.69ng/ml. The threshold were similar for Pathways 7 and 11 (Likert ≥ 3), For Pathways 9 and 13 (Likert ≥ 4),

using the maximum Youden method the optimal PSA threshold was 4.72ng/ml, to maintain sensitivity at 90% it was 0.9ng/ml and to maximise sensitivity it was 0.9ng/ml.

9.4.2 Secondary outcomes

A summary of the secondary performance outcomes for each pathway is shown in Table 46 and each performance metric will be explored in the following sections:

9.4.2.1 Biopsy Rates

A comparison between biopsy rates across all pathways is shown in Figure 53. The direction of the effect on biopsy rates is illustrated in left panel (A) and the p-values on the right panel. Pathways 1, 2, 3 & 7 had a high biopsy rate at 10.0% (95% CI 7.4-13.5), 17.7% (95% CI 7.4-13.5), 23.2% (95% CI 19.2-27.6) and 14.3% (95% CI 11.1-18.2) respectively. Pathways 4, 5 & 6 had similar rates to the reference standard. Pathways 8 to 13 had lower biopsy rates.

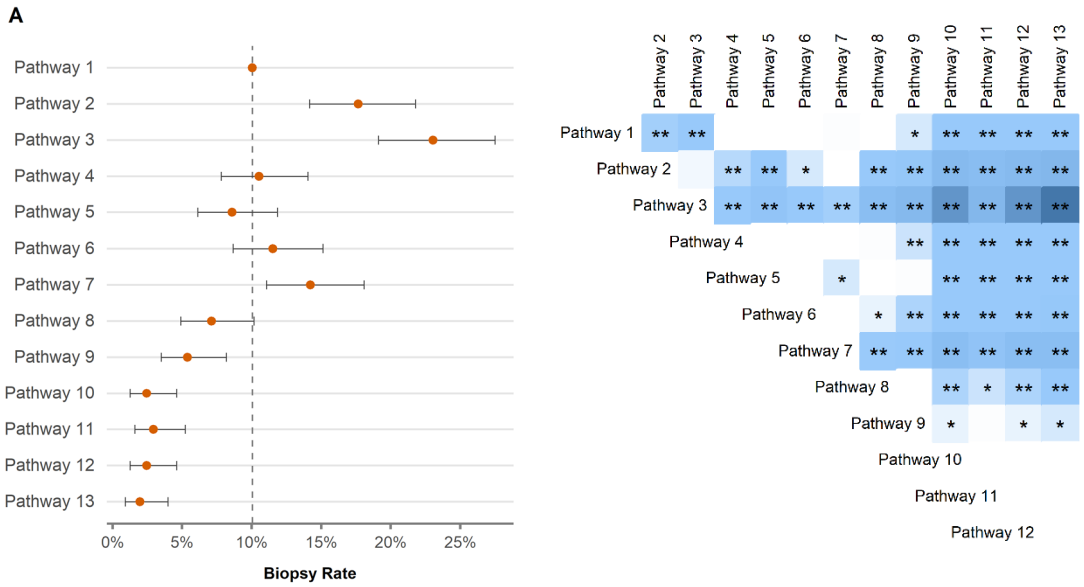


*Figure 50: Comparison of biopsy rates between pathways. (A) The proportion of participants biopsied in each pathway. Error bars represent 95% confidence intervals. (B) p-value for each pairwise comparison between pathways. The colour indicates where there is a statistically significant difference in proportion. A p<0.05 is indicated by a single *. If p <0.01 it is indicated by **.*

Table 46: Performance of each pathway

	Pathway 1	Pathway 2	Pathway 3	Pathway 4	Pathway 5	Pathway 6	Pathway 7	Pathway 8	Pathway 9	Pathway 10	Pathway 11	Pathway 12	Pathway 13
Tests	PSA ≥ 3	PIRADS ≥ 3	Likert ≥ 3	PIRADS ≥ 4	Likert ≥ 4	PSA ≥ 1 + PIRADS ≥ 3	PSA ≥ 1 + Likert ≥ 3	PSA ≥ 1 + PIRADS ≥ 4	PSA ≥ 1 + Likert ≥ 4	PSA ≥ 3 + PIRADS ≥ 3	PSA ≥ 3 + Likert ≥ 3	PSA ≥ 3 + PIRADS ≥ 4	PSA ≥ 3 + Likert ≥ 4
MRI rate*	0.0% (0-1.2)	100% (98.8-100)	100% (98.8-100)	100% (98.8-100)	100% (98.8-100)	46.8% (41.9-51.8)	46.8% (41.9-51.8)	46.8% (41.9-51.8)	46.8% (41.9-51.8)	10.0% (7.4-13.5)	10.0% (7.4-13.5)	10.0% (7.4-13.5)	10.0% (7.4-13.5)
Biopsy rate*	10.0% (7.4-13.5)	17.7% (14.2-21.9)	23.2% (19.2-27.6)	10.6% (7.9-14.1)	8.6% (6.2-11.9)	11.6% (8.7-15.2)	14.3% (11.1-18.2)	7.1% (4.9-10.2)	5.4% (3.5-8.2)	2.5% (1.3-4.6)	3.0% (1.6-5.2)	2.5% (1.3-4.6)	2.0% (0.9-4.0)
Detection rate*	1.0% (0.3-2.7)	3.2% (1.8-5.6)	3.4% (2.0-5.9)	2.7% (1.4-4.9)	2.0% (0.9-4.0)	2.7% (1.4-4.9)	3.0% (1.6-5.2)	2.5% (1.3-4.6)	1.7% (0.8-3.7)	1.5% (0.6-3.4)	1.5% (0.6-3.4)	1.5% (0.6-3.4)	1.2% (0.5-3.0)
Biopsy compliance	87.8% (6.3-95.4)	90.3% (12.7-95.7)	90.4% (17.1-95.3)	88.4% (6.8-95.6)	85.7% (5.1-94.6)	89.4% (7.6-96.0)	91.4% (10.0-96.8)	89.7% (4.3-97.3)	86.4% (2.9-96.4)	90.0% (1.1-99.5)	91.7% (1.4-99.6)	90.0% (1.1-99.5)	87.5% (0.8-99.3)
Overdiagnosis rate*	0.7% (0.2-2.3)	1.0% (0.3-2.7)	1.5% (0.6-3.4)	1.2% (0.5-3.0)	0.2% (0.0-1.6)	0.7% (0.2-2.3)	1.0% (0.3-2.7)	0.7% (0.2-2.3)	0.2% (0.0-1.6)	0.0% (0.0-1.2)	0.2% (0.0-1.6)	0.2% (0.0-1.6)	0.0% (0.0-1.2)
False positive rate	9.1% (6.5-12.4)	14.5% (11.3-18.4)	19.7% (16.0-24.0)	7.9% (5.5-11.1)	6.7% (4.5-9.6)	8.9% (6.4-12.2)	11.3% (8.5-14.9)	4.7% (2.9-7.3)	3.7% (2.2-6.2)	1.0% (0.3-2.7)	1.5% (0.6-3.4)	1.0% (0.3-2.7)	0.7% (0.2-2.3)
PPV	11.1% (3.6-27.0)	20.0% (11.5-32.1)	16.5% (9.6-26.4)	28.9% (16.0-46.1)	26.7% (13.0-46.2)	26.2% (14.4-42.3)	22.6% (12.7-36.5)	38.5% (20-59.3)	36.8% (17.2-61.4)	66.7% (30.9-91.0)	54.5% (24.6-81.9)	66.7% (30.9-91.0)	71.4% (30.3-94.9)
Intervention rate*	1.0% (0.3-2.7)	2.2% (1.1-4.3)	2.2% (1.1-4.3)	2.0% (0.9-4.0)	1.5% (0.6-3.4)	1.7% (0.8-3.7)	1.7% (0.8-3.7)	1.7% (0.8-3.7)	1.2% (0.5-3.0)	1.0% (0.3-2.7)	0.7% (0.2-2.3)	1.0% (0.3-2.7)	1.0% (0.3-2.7)
Number needed¥													
To screen	102	31	29	37	51	37	34	41	58	68	68	68	82
To biopsy	10	6	7	4	4	4	5	3	3	2	2	2	2

Data are n, n (%), mean (SD), and % (95% CI)

*Denominator is the number of eligible men for the pathway

¥To detect one participance with Gleason ≥ 3+4

9.4.2.2 Cancer Detection Rate

In total, 17 men were diagnosed with clinically significant prostate cancer (overall detection rate 4.2%, 95% CI 2.6-6.6). The cancer detection rates for each pathway are shown in Figure 54. The confident intervals were wide due to the low event rate but there was a consistent trend of decreasing cancer detection rates from Pathway 2 at 3.2% (95% CI 1.8-5.5) to Pathway 13 at 1.2% (95% CI 0.5-3.0).

9.4.2.3 Overdiagnosis Rate

During the study there were 20 men diagnosed with clinically insignificant prostate cancer (overall overdiagnosis rate 5.9%, 95% CI 3.2-7.5). The overdiagnosis rate appeared to be broadly similar across each pathway although similarly wide confident intervals were seen.

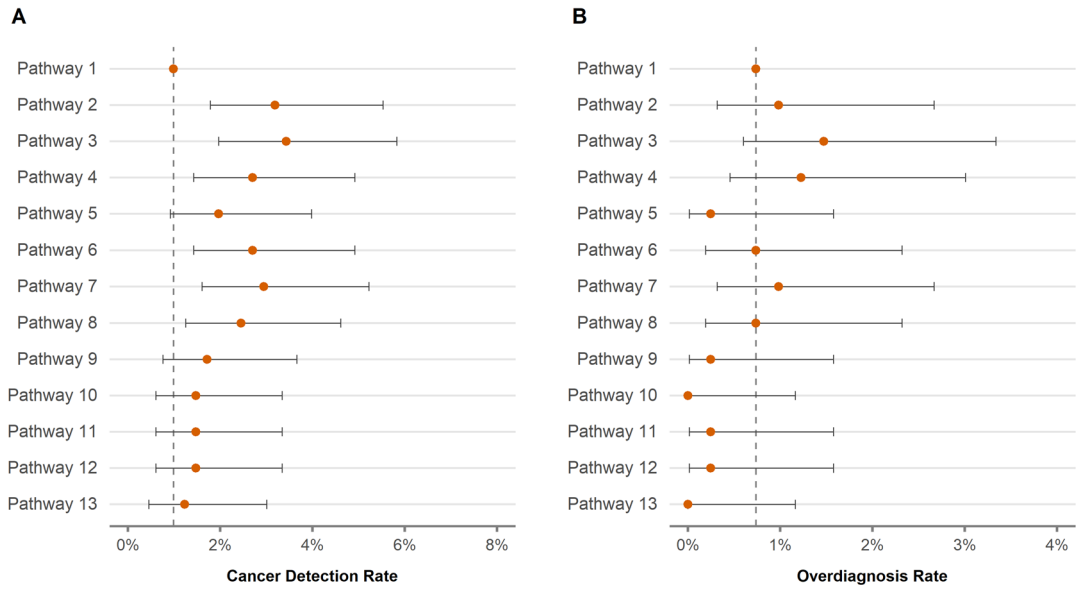


Figure 51: Comparison of cancer detection and overdiagnosis rates between each pathway. Error bars represent 95% confidence intervals. (B) p-value for each pairwise comparison.

Table 47: Change in Performance metrics in comparison to reference pathway

	Pathway 1 (reference)	Pathway 2	Pathway 3	Pathway 4	Pathway 5	Pathway 6	Pathway 7	Pathway 8	Pathway 9	Pathway 10	Pathway 11	Pathway 12	Pathway 13
Tests	PSA ≥ 3	PI-RADS ≥ 3	Likert ≥ 3	PI-RADS ≥ 4	Likert ≥ 4	PSA ≥ 1 + PI-RADS ≥ 3	PSA ≥ 1 + Likert ≥ 3	PSA ≥ 1 + PI-RADS ≥ 4	PSA ≥ 1 + Likert ≥ 4	PSA ≥ 3 + PI-RADS ≥ 3	PSA ≥ 3 + Likert ≥ 3	PSA ≥ 3 + PI-RADS ≥ 4	PSA ≥ 3 + Likert ≥ 4
Biopsy rate*	Ref	+7.6% (2.6-12.6)	+13.0% (7.7-18.3)	+0.5% (-3.9-4.9)	–1.5% (-5.7-2.8)	+1.5% (-3.0-6.0)	+4.2% (-0.5-8.9)	–2.9% (-7.0-1.1)	–4.7% (-8.6--0.8)	–7.6% (-11.1--4.1)	–7.1% (-10.7--3.5)	–7.6% (-11.1--4.1)	–8.1% (-11.5--4.6)
Biopsy compliance	Ref	+2.5% (-11.6-16.5)	+2.6% (-10.8-16.0)	+0.6% (-13.9-15.0)	–2.1% (-19.5-15.3)	+1.6% (-13.3-16.5)	+3.6% (-10.9-18.0)	+1.9% (-14.9-18.6)	–1.4% (-20.4-17.5)	+2.2% (-21.1-25.5)	+3.9% (-18.6-26.3)	+2.2% (-21.1-25.5)	–0.3% (-25.6-25.0)
Detection rate*	Ref	+2.2% (0.0-4.4)	+2.5% (0.2-4.7)	+1.7% (-0.4-3.8)	+1.0% (-0.9-2.9)	+1.7% (-0.4-3.8)	+2.0% (-0.2-4.1)	+1.5% (-0.6-3.5)	+0.7% (-1.1-2.6)	+0.5% (-1.3-2.2)	+0.5% (-1.3-2.2)	+0.5% (-1.3-2.2)	+0.2% (-1.4-1.9)
Overdiagnosis rate*	Ref	+0.2% (-1.3-1.8)	+0.7% (-0.9-2.4)	+0.5% (-1.1-2.1)	–0.5% (-1.7-0.7)	+0.0% (-1.2-1.2)	+0.2% (-1.3-1.8)	+0.0% (-1.2-1.2)	–0.5% (-1.7-0.7)	–0.7% (-1.8-0.3)	–0.5% (-1.7-0.7)	–0.5% (-1.7-0.7)	–0.7% (-1.8-0.3)
False positive rate	Ref	+5.4% (0.7-10.0)	+10.5% (5.5-15.5)	–1.2% (-5.3-2.8)	–2.5% (-6.4-1.5)	–0.2% (-4.4-3.9)	+2.2% (-2.2-6.6)	–4.4% (-8.1--0.7)	–5.4% (-9.0--1.8)	–8.1% (-11.3--4.9)	–7.6% (-10.9--4.3)	–8.1% (-11.3--4.9)	–8.3% (-11.5--5.2)
PPV	Ref	+8.9% (-7.4-25.2)	+5.4% (-9.6-20.3)	+17.8% (-2.6-38.2)	+15.6% (-6.4-37.5)	+15.1% (-4.3-34.5)	+11.5% (-6.0-29.1)	+27.4% (2.7-52.0)	+25.7% (-2.3-53.7)	+55.6% (16.1-95.0)	+43.4% (6.3-80.5)	+55.6% (16.1-95.0)	+60.3% (16.8-100)
Intervention rate*	Ref	+1.2% (-0.7-3.2)	+1.2% (-0.7-3.2)	+1.0% (-0.9-2.9)	+0.5% (-1.3-2.2)	+0.7% (-1.1-2.6)	+0.7% (-1.1-2.6)	+0.7% (-1.1-2.6)	+0.2% (-1.4-1.9)	+0.0% (-1.4-1.4)	–0.2% (-1.8-1.3)	+0.0% (-1.4-1.4)	+0.0% (-1.4-1.4)

Data are n, n (%), mean (SD), and % (95% CI)

Abbreviations: TBx = Targeted Biopsy, N-TBx = Non-targeted biopsy, CBx = Combined Biopsy

*Denominator is the number of eligible men for the pathway

¥To detect one participance with Gleason ≥ 3+4

9.4.2.4 False positive rate

There were 150 participants who had a screen-positive result which did not lead to a diagnosis of significant prostate cancer. Comparison of the false positive rates between each pathway is illustrated in Figure 55. Pathways 1 and 2 had high false positive rates, pathways 4 to 7 had similar rates to the standard PSA pathway. Pathways 8 to 13 had low false positive rates.

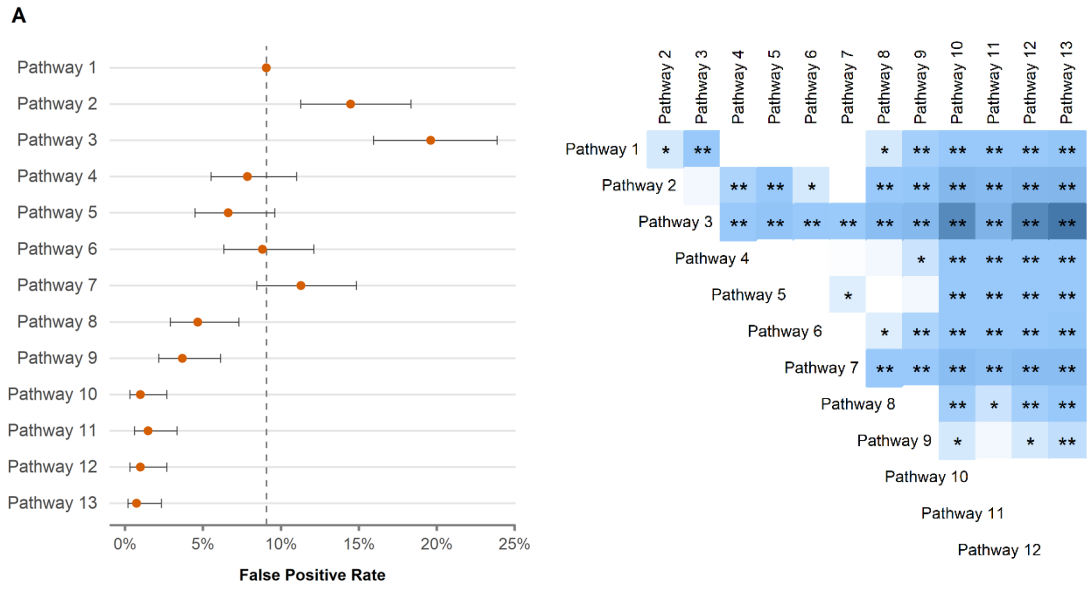


*Figure 52: Comparison of false positive rates between pathways. (A) The proportion of participants with a false positive result who underwent a biopsy. Error bars represent 95% confidence intervals. (B) p-value for each pairwise comparison. The colour indicates where there is a statistically significant difference in proportion. A p<0.05 is indicated by a single *. If p <0.01 it is indicated by **.*

9.4.2.5 Positive Predictive Value

PPV is calculated on men diagnosed with significant prostate cancer by those who underwent a biopsy. Similar to cancer detection rate there was a consistent increase in PPV across the majority of pathways as shown in Figure 56.

There was a sequential increase from 20.0% (95% CI 11.5-32.1) for Pathway 2 to 71.4% (95% CI 30.3-94.9) for Pathway 13.

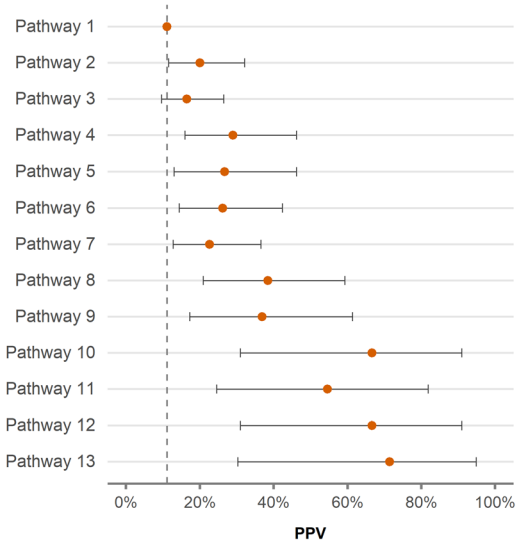


Figure 53: Positive Predictive Value for each pathway

9.4.2.6 Intervention Rate

The intervention rate was defined as the proportion of participants who underwent active and curative treatment for prostate cancer. This included nine patients who underwent either focal therapy (n = 4), radical prostatectomy (n=2) or radiotherapy (n=3).

A comparison between pathways is shown in Figure 57. Due to the low event rate the confidence interval were wide.

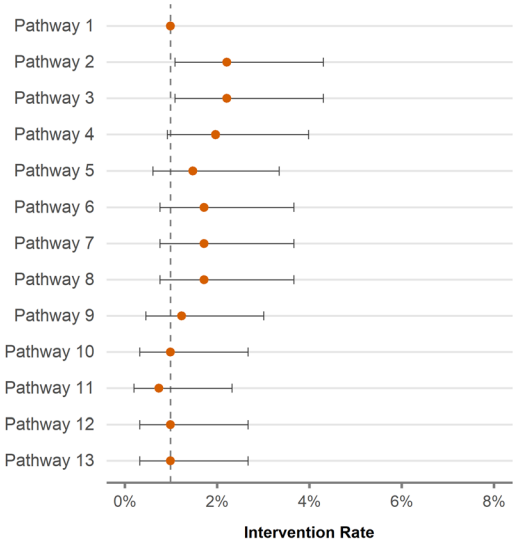


Figure 54: Intervention Rate for each pathway

9.5 Discussion

9.5.1 Principle findings

The design of the IP1-PROSTAGRAM trial provided a unique opportunity to assess the clinical performance of a variety of combinations of PSA and fast MRI in men undergoing prostate cancer screening. Each screening pathway was evaluated with respect to biopsy referral rates, false positives and disease detection. The primary analyses demonstrated an explicit trade-off between biopsy rates versus significant cancer detection..

The findings show that pathways which incorporated PSA as a triage test (Pathways 8-13) would recommend fewer men for biopsy with low false positive rates. These combined strategies were more efficient compared to the single test strategies (Pathways 1-5) as indicated by the lower number of biopsies needed per cases of significant cancer detected. We found low efficacy for the standard screening pathway (Pathway 1) with large numbers of biopsies performed with few cases of significant prostate cancer.

The standard PSA pathway (Pathway 1) was shown to have a 10.0% biopsy rate and false positive rate of 9.1%. This standard pathway generates a substantial burden for men and has implications for healthcare systems given the high costs and risks of biopsy. Pathways 2-3 which relied on fast MRI alone led to higher biopsy and false positive rates but with a benefit of high cancer detection rates. These data support the findings of previous chapters that at

PSA ≥ 3ng/ml, there are a number of patients with significant prostate cancer which are overlooked. Comparison to previous studies

The outcomes can be compared to those reported by the pilot Göteborg 2 screening study[311]. In this study men were randomised to three screening pathways of PSA≥ 3ng/ml (equivalent to Pathway 1), PSA≥ 3ng/ml + PI-RADS ≥ 3 (equivalent to Pathway 10) and PSA ≥ 1.8 and PI-RADS ≥ 3 (similar to Pathway 6). This study found a similar reduction in biopsy rate with the addition of MRI to PSA ≥ 3ng/ml. The proportion of men biopsied in each arm was different to the results of this chapter which is expected as the study included men within the 10th screening round of the Göteborg randomised screening trial who had already undergone multiple rounds of screening.

As the optimal threshold for PSA in association with a screening MRI has not been previously investigated, the Göteborg 2 trial chose the cut-off 1.8ng/ml arbitrarily[311]. In this study we evaluated pathways 6-9 with a lower PSA threshold at 1ng/ml and evaluated the optimal cut-off for PSA across different MRI scores. The findings suggest that a threshold around 1ng/ml would provide an appropriate balance between reducing biopsy burden without impacting detection of significant prostate cancer.

This is supported by longitudinal studies which have shown that men with a PSA less than 1.0ng/ml have a low lifetime risk of prostate cancer specific mortality[310]. A pathway with PSA ≥ 1.0ng/ml alone is not feasible as it would lead to a biopsy rate of 46% in this study and a sequential MRI is needed to reduce the biopsy rate. Similarly, a number of the MRI-alone pathways would lead to a higher number of biopsies compared to the standard pathway.

9.5.2 Implications of findings

The primary outcome of this chapter demonstrates that a trade-off exists between biopsy rates and cancer detection rates across various pathways (Figure 51). Policy makers and clinicians will have to determine the optimum trade-off given the competing risks. In this decision there is an inevitable trade-off between maximising cancer detection rates and minimising reducing biopsy rates. Pathways which have the highest detection rates have correspondingly high biopsy rates which will be associated with biopsy-related morbidity and overdiagnosis.

The pathways which combine PSA and MRI appear to offer the most appropriate balance between the risks of biopsy, false positives and false negatives when compared to the current

pathway (Pathway 1). The current screening pathway using PSA alone (Pathway 1) recommended more patients for biopsy and is less efficient in terms of the number needed to screen or biopsy. The number needed to biopsy to diagnose a significant prostate cancer was 10 men for Pathway 1 compared to 3 men for Pathway 8. In the UK, scaling up the results from this analysis suggests that if Pathway 1 and Pathway 8 were implemented as screening tests and there was 100% compliance, Pathway 8 would result in 227,922 men needing a biopsy and avoid 354,814 men receiving a false positive result.

Further, it is worth noting that an effective screening pathway for prostate cancer requires a high compliance with biopsy recommendations. It is encouraging that the compliance rate was between 87.8% and 91.4% across each pathway after a positive result. In other screening programmes the accepted quality indicator for adherence is 90% after a positive screening test[312] and therefore all the pathways meet this recommendation.

9.5.3 Limitations

This chapter was subject to several limitations.

First, it is acknowledged that the multiple comparisons between 13 pathways could lead to certain factors being falsely significant (Type 1 error). Methods to correct for multiple hypothesis testing, such as Bonferroni adjustment, were not employed due to the correlation between the pathways. Bonferroni adjustment and other methods require an assumption of test independence which was not met as the pathways are highly related due to the similarities in tests.

Overcorrecting for a Type 1 error in the setting of highly correlated tests increases the risk of Type 2 error which occurs when significant inferences are missed. As an example, it is expected that Pathway 2 and Pathway 3 will perform similarly as differences in MRI scoring system is likely to be minimal[313]. However, applying Bonferroni adjustment leads to differences in biopsy rates which leads to Pathway 2 becoming insignificant with respect to Pathway 1 ($p = 0.00236$ without correction, $p = 0.18421$ with correction) while Pathway 3 remains significant with respect to Pathway 1 despite correction. This is counter intuitive as the similarities between Pathways 2 and Pathway 3 in terms of MRI scoring systems re-enforce rather than detract from the conclusion that an MRI at score ≥ 3 leads to an increase in biopsy rate.

It is accepted that under certain clinical decision scenarios, formal adjustment of multiple comparisons is not necessary[314]. This is supported by the fact that fast MRI as a screening test is at an exploratory stage and it is common during the early stage of pathway development to start with a large number of potential pathways which can be selected for further validation in a larger cohort. Therefore, the findings of this chapter will pave the way for further studies validating the MRI pathway.

Second, the comparison of cancer detection rates, overdiagnosis rates and intervention rates was limited due to the low event rates. There were 17 patients with significant prostate cancer, 20 patients with insignificant prostate cancer and 9 patients who underwent active treatment. Due to the low event rate it was not feasible to evaluate whether there was a significant difference in outcomes for these performance metrics across each pathway.

Third, the fast MRIs were reported by radiologists blinded to PSA. This does not reflect the realities of a combined pathway where radiologists would have access to the PSA result at the time of reporting. A recent meta-analysis has shown that PSA combined with prostate volume (PSA density) may be a useful factor to predict clinically significant prostate cancer in men with a non-suspicious MRI[315]. Although this was not evaluated in this study it is possible that a combined PSA and MRI pathway may have further improvements in performance in this scenario if radiologists are unblinded to clinical information such as PSA.

Fourth, all pathways apart from the reference pathway (Pathway 1) utilised MRI-targeted biopsy alone to detect significant disease. Currently an MRI-targeted and systematic biopsy is considered the optimal approach to diagnose men with a MRI visible lesion and MRI targeted biopsy-alone has not been considered sufficiently accurate to safely replace a combined targeted and systematic biopsy[316]. This combined approach has been shown to have a high detection rate for significant prostate cancer and reduce grade misclassification[316]. However, due to multiple tests used in IP1-PROSTAGRAM it increases the reliability of results to only include cores from the MRI targets.

9.6 Conclusion

This chapter has evaluated the performance of different combinations of PSA and fast MRI to determine which men undergoing screening should be recommended for a biopsy. It has highlighted the trade off which exists between reducing excessive numbers of biopsy and false positives while maintaining cancer detection rates. These outcomes will help policy-

makers to navigate the various combinations of PSA and MRI selecting the optimal new screening pathway.

Chapter 10 – Evaluating a combined PSA and MRI pathway and set up of a multi-centre registry

10.1 Overview

This chapter evaluates a potential pathway which could deliver a screening programme which combines PSA and an MRI. It utilises data from a national registry which I designed, implemented and maintained during the research period. The chapter includes my learnings from setting up this national registry in the context of improving the overall benefit:risk profile for screening and diagnosis of prostate cancer. A portion of this chapter forms the basis of work I presented at BAUS 2019[ix].

10.2 Introduction

Giving the findings of the previous chapter that a new screening pathway may combine PSA and a fast MRI, it is important to evaluate possible methods for delivering such a pathway. Setting up an organised screening pathway on a large-scale will present a significant challenge for healthcare systems given the practicalities of delivering high-volume, imaging-based prostate cancer care.

Even delivering pre-biopsy diagnostic MRI has been a challenge due to the lack of availability of MRI. Current guidelines recommend that men who undergo opportunistic PSA screening with a positive result have a multiparametric MRI [317] but this can cause delays. One solution which has been piloted which could be applicable to a screening setting is the Rapid Access Prostate Imaging and Diagnosis (RAPID) pathway.

The RAPID pathway was commissioned by NHS England to evaluate whether it could be an effective model for delivering high-volume prostate cancer diagnostics. It represents a new approach to delivering MRI directed diagnostics in a sustainable, high quality and timely manner. The hallmark of this pathway was a one-stop model similar to the approach offered following breast cancer screening. It addresses many of the problems with the conventional pathway for prostate cancer which requires a lengthy stepwise series of tests and multiple hospital appointments.

[ix] Eldred-Evans, D., et al. (2019). " Rapid Access Prostate Imaging and Diagnosis (RAPID) pathway – an innovative approach for prostate cancer diagnosis." Journal of Clinical Urology Vol 12(1S): e36-37

This chapter evaluates whether this accelerated RAPID pathway could be an effective model for delivering high-volume prostate cancer screening. In order to evaluate the outcomes of this pathway a prospective diagnostic registry was designed. The RAPID Online registry was developed as a large, interoperable, national registry to report the longitudinal outcomes of all men in the RAPID pathway. The main advantage of this method of data collection was the ability to standardise data elements across each institution delivering the RAPID pathway. Further details regarding setting up the registry are provided in the methods section.

The registry can be used to monitor the performance of the pathway in accordance with the 2019 NHS Long Term Plan which committed to improving cancer outcomes and reducing variations in care with faster diagnostic standards and more streamlined patient pathways[318]. Although there remains controversy over the impact of such performance targets on cancer outcomes, a timely diagnosis has been highlighted as an effective means to decrease psychological distress and increase patient satisfaction following a referral with a suspicion of cancer[319]. Streamlined pathways can also increase capacity by reducing unnecessary appointments and tests. It is common for healthcare systems to use diagnostic interval as an indicator for quality of cancer care

RAPID was set up address the challenge around waiting times and inter-site variability in the prostate cancer diagnostic pathway. The aim was to reduce diagnostic delays, increase quality through multidisciplinary team decisions and increase patient wellbeing and satisfaction. Similar strategies have been reported for low-dose CT for lung cancer[320] and colonoscopy for bowel cancer[321] but there has been limited investigations into mpMRI as a straight-to-test strategy.

In this chapter, the impact of the RAPID pathway was evaluated by comparing diagnostic interval using an interrupted time series. The null hypothesis was that the median time to diagnosis pre-and post introduction of the RAPID pathway were similar.

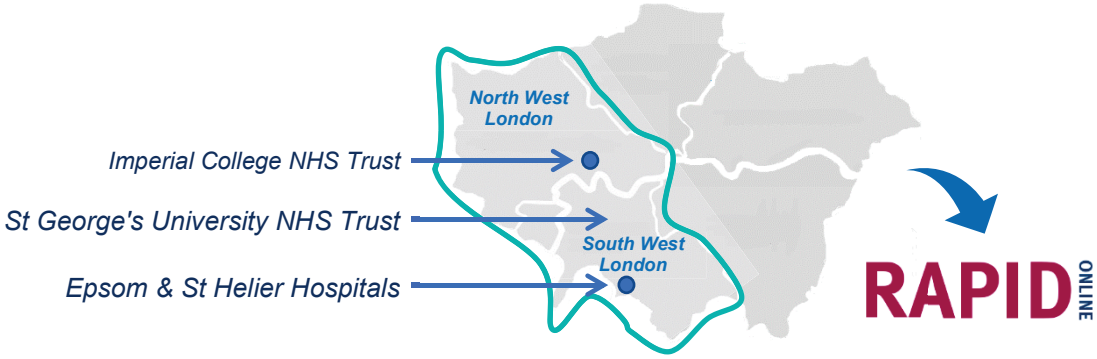
10.3 Methods

10.3.1 Study Design

This was a mixed study which included an interrupted time-series (ITS) analysis to evaluate the impact of the RAPID pathway on diagnostic interval. ITS is considered one of the strongest quasi-experimental designs for evaluating the impact of an intervention where longitudinal data has been collected before and after commencing an intervention [322]. The ITS analysis was reported against the Quality Criteria for Interrupted time series design [323]. Primary outcomes from the ITS analysis were supplemented by clinical outcomes collected in RAPID Online and patient satisfaction surveys.

10.3.2 Study Setting

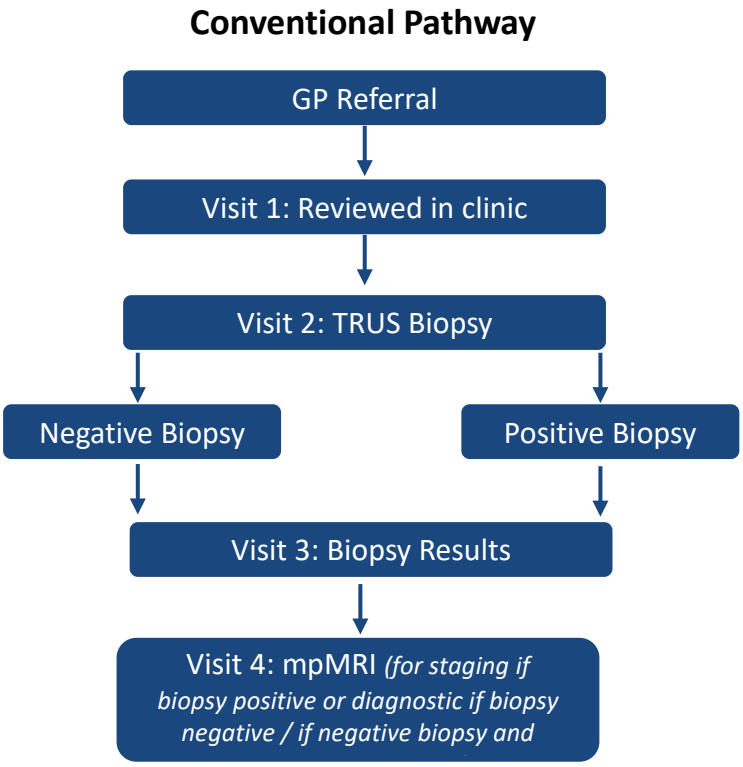
The RAPID pathway was set-up at three hospitals (two university and one general hospital) in the United Kingdom. These hospitals provide secondary and tertiary level prostate cancer services to a population of 3.8 million people in London. There was a stepwise introduction of the RAPID pathway across sites with Imperial College Healthcare NHS Trust commencing in May 2017, Epsom and St Helier Hospitals in November 2017 and St George's University Hospitals NHS Foundation Trust in March 2018.



10.3.3 The intervention

In the convention pathway, Visit 1 required men to have an outpatient appointment with additional simple tests such as repeat PSA and urine cultures, Visit 2 was a transrectal ultrasound-guided prostate biopsy (TRUS-biopsy) and Visit 3 required to provide the results. Often a 4th visit had to be scheduled for a staging MRI or diagnostic MRI if there was continual concern following a negative TRUS biopsy. This fourth visit had to be scheduled four weeks

after the TRUS biopsy to minimize post-biopsy artefact which degrades the performance of mpMRI[324, 325].



The RAPID pathway was set up with the objective of providing men with the option to undergo all investigations to rule in or out prostate cancer during a single visit. Following the introduction of the RAPID pathway all eligible men underwent a mpMRI as the first-line investigation prior to a prostate biopsy. The eligibility criteria for RAPID required patients to be clinically appropriate for mpMRI +/- transperineal prostate biopsy. Referrals were vetted to exclude men who would have not met predefined criteria for RAPID due to age, co-morbidities or inability to undergo all investigations in RAPID.

All mpMRIs were acquired in accordance to standards set out in the guidelines[201, 326] and scored on an ordinal scale from 1 to 5 based on the likelihood of significant prostate cancer (1 = very low, 2 = low, 3 = indeterminate, 4 = high, 5 = very high). A suspicious mpMRI was defined as an MRI Score >/= 4 or a Score 3 with a high PSA density (>/=0.12). The high NPV of mpMRI allowed the pathway to decrease the number of prostate biopsies and move to a more targeted approach. So men with a non-suspicious MRI were deemed at low risk of significant cancer and discharged from the pathway in accordance to standards set out in NICE guidelines [317].

The mpMRIs were performed using a 1.5T scanner with a pelvic phased-array coil in accordance with the European Society of Uroradiology guidelines[208]. The MRI scanner, radiology experience and reporting methods varied between sites as summarised in Table 48. The standard MRI sequences included T1-weighted, T2-weighted, multi b-values (for derivation of apparent diffusion coefficient (ADC) maps with a high b-value of 1500. Dynamic contrast-enhanced (DCE) sequences were taken with two pre-contrast and thirteen postcontrast dynamic series after intravenous administration of 0.1mmol/kg gadolinium-based contrast material and flushing with 20mL saline. A standard single dose of intravenous buscopan was administered prior to image acquisition and no endorectal coil was used. Slice thicknesses were: axial T2WI- 3 mm, axial T1WI- 5 mm, coronal T2WI- 4mm, DWI- 4mm, DCE-3mm.

Table 48: MRI and reporting across RAPID sites

Site	**MRI scanner**	**MRI protocol**	**MRI Reporting**	**Radiologists experience**
Site 1	1.5T Siemens Aera + pelvic phased array coil	Axial, coronal, sagittal T2w; Axial T1w; DWI, DCE-MRI	PI-RADs version 2.0#	2 radiologists with 8 and 10 years' experience
Site 2	1.5T Siemens Avanto + pelvic phased array coil	Axial, coronal, sagittal T2w; Axial T1w; DWI, DCE-MRI	PI-RADs version 2.0# and Likert scoring	4 radiologists with 9, 6, 5 and 5 years experience
Site 3	1.5T Siemens Avanto + pelvic phased array coil	Axial, coronal, sagittal T2w; Axial T1w; DWI, DCE-MRI	PI-RADs version 2.0# and/or Likert scoring	2 radiologists with > 5 years experience

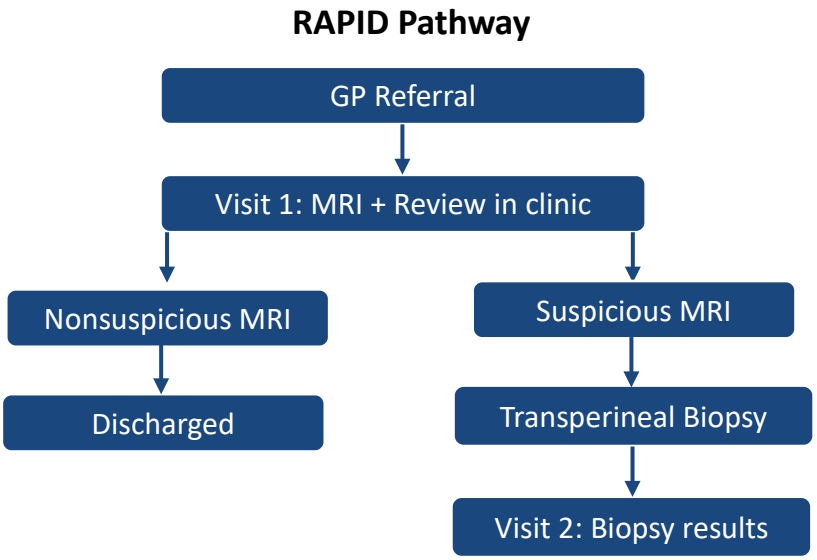
Following the MRI, the aim was to provide same day reporting so patients could be immediately advised on their mpMRI result prior to leaving the hospital. Patients had a face-to-face consultation with a clinician who informed them of the MRI result and advise whether further investigation with a prostate biopsy was recommended to reach a diagnosis. During this clinic those patients with a non-suspicious MRI were discharged from the pathway and advised on any need for ongoing PSA monitoring. The general practitioner (GP) received a letter to this effect which included an individualised PSA level warranting re-referral on a prostate cancer pathway.

Men with a suspicious mpMRI were advised to have a prostate biopsy which could be performed during the same visit depending on patient preference and site-specific biopsy availability. All biopsies were performed via the transperineal route to minimize risk of biopsy-related morbidity, particularly sepsis. The procedure occurred in either day surgery or the outpatient setting. Software assisted registration was available using Biopsee mpMRI

Fusion Biopsy System (MedCom GmbH, Darmstadt, Germany). The biopsy protocol included sampling of both MRI-targeted and contralateral systematic biopsy of the non-suspicious gland.

Biopsies were performed by specialist nurses or urologists depending on site preference. Certain sites offered local anaesthetic biopsy which could be performed in an outpatient setting. Targeted transperineal prostate biopsies were performed using either visual-estimation or image-fusion under local anaesthesia (LA), combination sedation and LA, or general anaesthesia (GA). During all procedures, a biplanar transrectal ultrasound probe was used (Hitachi, Japan). Image-fusion targeting was performed using the BiopSee® platform (Medcom, Germany). To perform image-fusion registration, operators imported MRI images onto the BiopSee® system, contoured the prostate and all lesions to be targeted and then fused these contours with transrectal ultrasound images acquired using a manual pull-back on an electronically tracked stepper.

For ongoing quality assurance of reporting, mpMRIs (suspicious and non-suspicious) with targeted biopsy results were re-reviewed in a multi-disciplinary team (MDT) meeting with a urologist, radiologist and histopathologist. If prostate cancer was found on biopsy the patient's care was taken over to the hospital's multi disciplinary team for ongoing prostate cancer management



10.3.4 The RAPID registry

Implementing the RAPID Online Registry required the following steps:

Defining and selecting data elements: The RAPID consortium agreed the common data standards and terminology to be used in the database. These were agreed during a meeting including urologists, radiologists and nurses who would be delivering the RAPID pathway. A standard operating procedure for data entry and registry workflow was prepared to ensure that data collection activities were feasible for each site. During annual pathway meeting, the data points were refined and standardized definitions agreed.

Development of electronic case report forms: Following agreement of the standardized data entry elements case report form (CRFs) were developed using the REDCap (Research Electronic Data Capture) electronic data capture tool hosted at Imperial College London. Data entry was designed so that collection was aligned to clinical workflow in order to maximise efficiency. REDCap is a web-based data management tool which allows data-entry across multiple sites. Additional characteristics include (1) a simple interface to allow consistent data entry with automated data validation checks. Figure 58 shows an example of a case report form (2) detailed audit for tracking user data entry (3) automated export procedures to statistical packages such as R studio and (4) anonymisation to maintain data confidentially according to GCP guidelines and the Health Insurance Portability and Accountability Act (HIPAA).

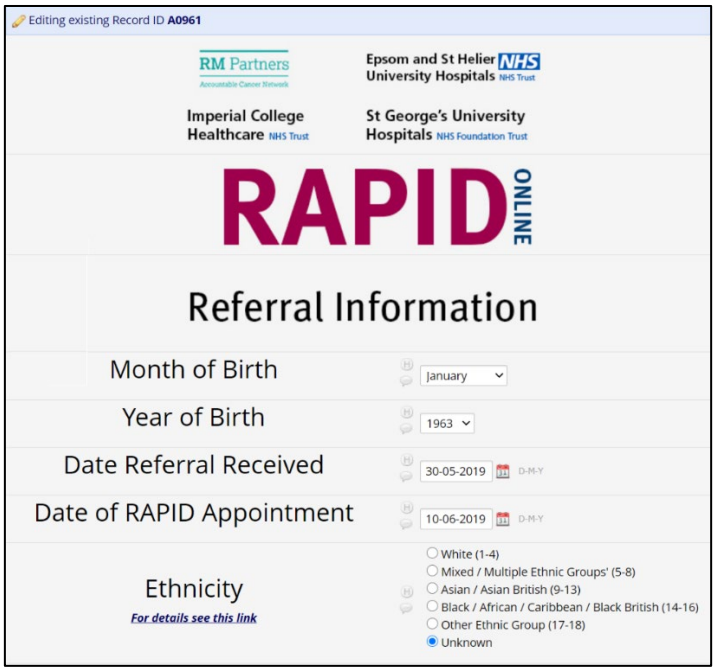


Figure 55: An electronic case report form (CRF) in RAPID Online Data.

Training of staff for data collection: All users went through semi-structured training prior to commencing data entry. The goal was to give users an overview of RAPID Online whilst also making sure that any factors that could compromise the integrity of data collection were identified. There was a continual optimisation process to take account of any issues highlighted during training. Post training monitoring of the data entry was carried out for the first 20 cases to ensure data quality. Where there were data inconsistencies this was fed back to users and where necessary retraining was implemented.

Quality control of data: A system was developed to generate data quality control reports using R statistical software. The reports were transformed into a csv file and distributed to the database coordinator. Quality control at a site level was ensured by requiring users to enter data at a pre-specified time-point, three months after referral, to ensure each patient had finished the pathway for complete data entry. Monitoring of the data collection was conducted and source data verification was conducted where necessary.

Augmenting with additional data: Data collection on cancer waiting times is mandated by the NHS and recorded prospectively by each hospital. Each RAPID hospital has a dedicated database for reporting to the National Cancer Waiting Times Monitoring Data Set. Data was extracted for 24 months prior to initiation of the RAPID pathway (pre-intervention period) and for 24 months following the commencement of the pathway (intervention period). We identified consecutive patients referred through the two-week wait referral pathway for suspected prostate cancer following the conventional pathway in the pre-interventional period and those following the RAPID pathway in the interventional period. Selection criteria for the two-week-wait pathway follows national guidance from NICE [317]. To ensure data quality, reliability of data extracted was assessed on a 10% sample of waiting times. Missing data was supplemented with information from additional hospital data systems where necessary.

Additional acceptability data: For evaluation of patient acceptability, a written patient satisfaction survey was developed by RM Partners West London Cancer Alliance in conjunction with a patient advisory group. An independent research company Ipsos MORI (London, UK) was commissioned to conduct additional qualitative interviews from a sample of patients. Patients were encouraged to complete their questionnaires and return them to RM Partners so the results could be evaluated independent of sites staff. The questionnaires recorded basic demographic data and comprised 16 questions formulated as a series of

statements focused on the acceptability of an accelerated diagnostic pathway for prostate cancer.

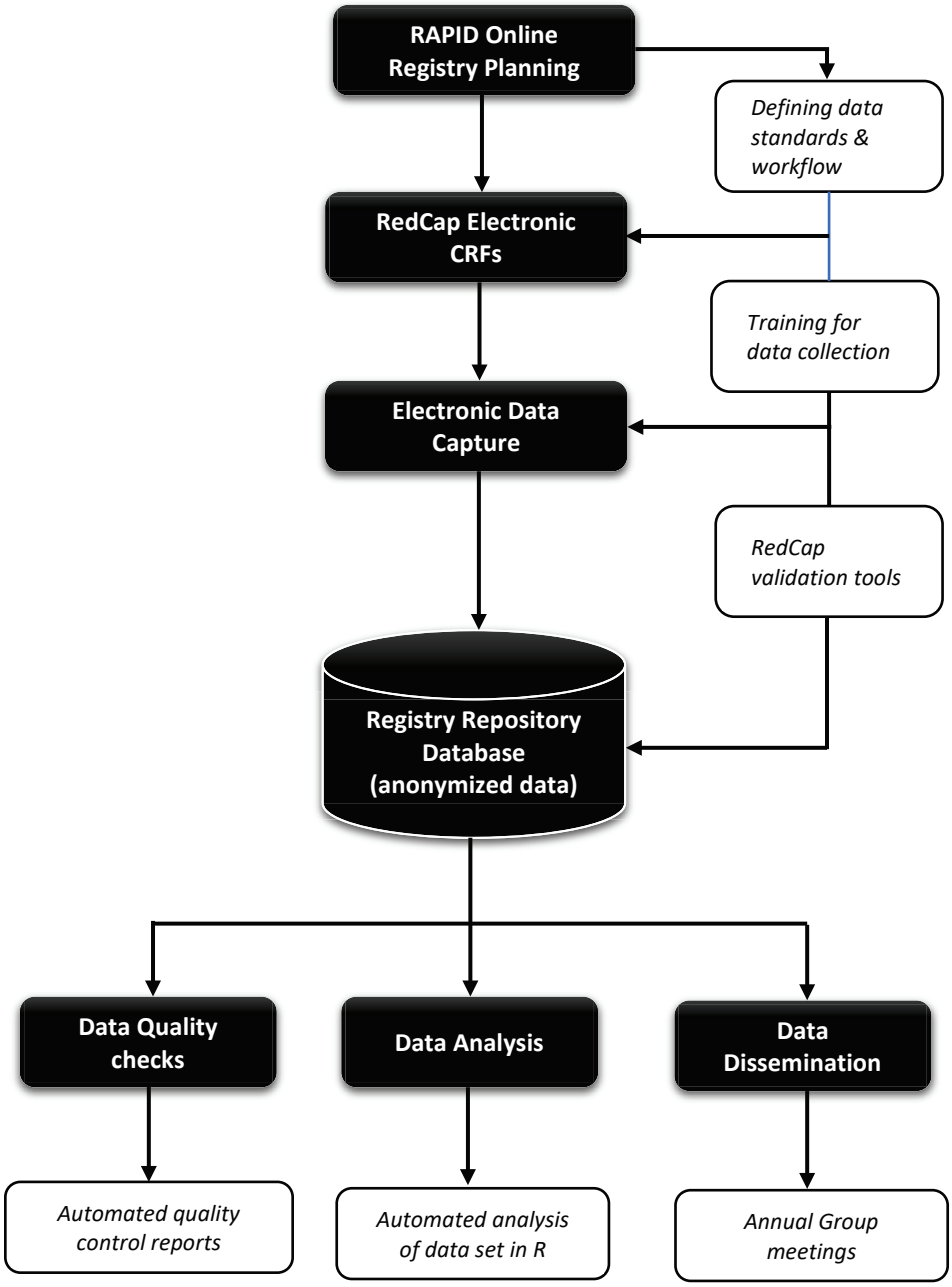


Figure 56: Process of RAPID Registry.

10.3.5 Outcomes

The primary outcome of this study was the change in median time to diagnosis between the pre-intervention and intervention period. Time to diagnosis was defined as the number of days between receipt of referral and diagnosis date. For consistency between sites, diagnosis date was determined using MDT date in men who underwent biopsy (histological diagnosis), MRI report date for men who did not require a biopsy (radiological diagnosis) and clinic date for clinically diagnosed cases requiring neither radiological nor histological diagnosis.

Secondary outcomes included the proportion of men who avoided a biopsy and detection of prostate cancer stratified by mpMRI score. A site-level comparison of clinical data included radiological outcomes, biopsy techniques, biopsy complications and patient satisfaction survey results.

10.3.6 Statistical Analysis

Primary outcome: This was assessed using segmented linear regression which divides a time series into pre- and postintervention segments. ITS is a quasi-experimental design which is commonly used to evaluate the impact of health-care interventions where data has been collected prior to and after introduction[327-329]. ITS are particularly useful in situations where RCTs are impractical, unethical or not cost-effective. In contrast to other pre- and post techniques, such as Shewhart control charts, ITS allows secular effects to be controlled and reduces the chance that the observed differences are caused by a pre-existing trend rather than the intervention. ITS is also able to detect changes in the trend such as a continual reduction in time to diagnosis following introduction of a new pathway.

For the ITS analysis, we constructed a time series of monthly median days to diagnosis aggregated across all sites using a segmented regression model. The data was divided into two groups (1) 24 months prior start of RAPID (pre-intervention period) and (2) 24 months following the commencement of the pathway (intervention period). The analysis included 24 datapoints before and after the start date of the RAPID pathway (48 in total). Monthly intervals were selected to allow evaluation of seasonal variation while ensuring sufficient observations at each time point [329]. As the start date of the pathway varied, the datapoints were intervention centred on the start date of the pathway at each site.

The general model took the form of

$$Y_t = \beta_0 + \beta_{1.}Time_t + \beta_{2.}level_t + \beta_{3.}trend_t + \varepsilon_t$$

Where

- Y_t is the primary outcome measure of monthly time to diagnosis
- β_0 represents the baseline time to diagnosis at the beginning of the study period
- β_1 is the baseline trend prior to the introduction of RAPID
- β_2 is increase or decrease in the level immediately after introduction of RAPID
- β_3 is the change in the trend in the rate of diagnosis after introduction of RAPID
- ε is a normally distributed random error

The model allowed an evaluation of the change in levels and trends of median time to diagnosis before and after introduction of the RAPID pathway. This was determined using four variables within the ITS reported in the analysis: baseline, baseline trend, level change and trend change. A statistically significant effect estimate for B_2 (level change) would suggest an immediate effect on waiting times while a statistically significant B_3 suggest a change in the trend over time from introducing RAPID.

ITS Procedure: All analyses for ITS were conducted in R Version 4.0.2 [213] using RStudio Version 1.0.44 and the NLME (Non-linear mixed effects) package. The procedure followed the steps by Wagner et al[330] which required the following steps

1. Stationarity: First, the unit root of median time to diagnosis using the Augmented Dickey Fuller test (ADF) was tested. This tests a key assumption of the ITS model which is that time series have stationarity. The auto-correlation plots and ADF test showed that the median time to diagnosis was not stationary (Figure 60B). Due to the presence of non-stationarity the series required differentiation meaning the difference of the median time to diagnosis from one month to the next was taken and then analysis of this differentiated series was conducted. The time series was differentiated in the first order and tested again to confirm stationary. The impact of this first order differentiation on the time series plot is shown in Figure 60C. The ACF plots and ADF test were repeated on the first order differentiation terms. The ACF

plots indicated that the outcome had became stationary after the first differentiation. The results of the ADF test showed that the null hypothesis of non-stationarity could be rejected.

2. Autocorrelation: Data involving time-series are often autocorrelated meaning that the events closer together in a time series tend to be more similar than events further apart in time). This makes the model residuals non-independent which is a key assumption for ordinary least-squares regression. If the residuals are not independent it can lead to under or over-estimate of the outcomes. Examination of the partial autocorrelation function for the centred dataset showed that no seasonality adjustment was required.

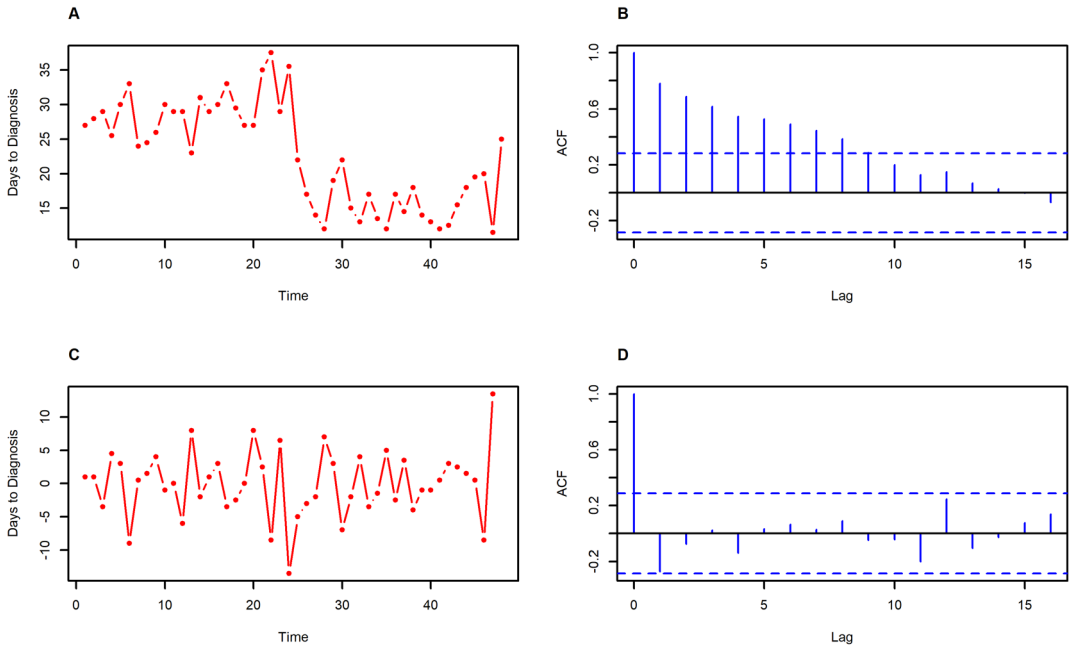


Figure 57: Properties of the raw monthly median days to diagnosis to assess for non-stationarity : (A) Time series plot of days to diagnosis suggesting an intercept and negative trend over the time series. (B) Auto-correlation (ACF) plot of raw data confirming negative trend (C) Time Series plot of First Order Differentiation of raw median time to diagnosis suggesting that first order differentiated data is stationary (D) ACF plot indicating a stable time series following first order differentiating.

Secondary outcomes: Demographic, radiological and histological characteristics across each RAPID site were presented as median (first and third quartiles; Q1 and Q3) or as number and percentage (%). Chi-squared test of independence was used to compare the distribution of categorical variables and the t-test was used for comparison of continuous variables. For the crude analysis, the time to diagnosis on a patient level before and after implementing the RAPID pathway was calculated using median days to diagnosis with IQR.

In order to assess variability between RAPID centres, the positive predictive values for clinically significant cancer according to PSA density and MRI suspicion levels for each site were compared. This was presented graphically using contour plots. All calculations were conducted in R Version 4.0.03.

10.3.7 Ethical considerations

This chapter evaluated a change in delivery of standard of care and so met the definition of a service evaluation under the NHS Health research authority guidelines. As such, ethical approval was not required and requirements for formal ethical approval were waived by the UK NHS Health Research Authority[331]. As all outcome measures are collected as part of routine care, the need for consent was waived. Institutional approval was obtained to collect anonymized data obtained during routine clinical practice. Each site has been designated as Site 1, Site 2 and Site 3 to preserve anonymity.

10.4 Results

10.4.1 Primary outcome

The study population included 5,565 patients referred with a suspicion of prostate cancer between 13 April 2015 and 31 March 2020. There were 3,435 during the pre-intervention period and 2,130 in the post-intervention period. For the primary endpoint, the ITS analysis showed an immediate and sustained change in time to diagnosis with no change in trend between pre- and post-intervention periods (Figure 61).

The estimated effect of the RAPID pathway was a reduction of 16.25 days (95% CI 12.13-20.37, p<0.001) in time to diagnosis. Time to diagnosis at end of the pre-intervention period was 32.1 days (95% CI 29.3-34.9) compared to 15.9 days (95% CI 12.9-34.9) during the intervention period representing an overall decrease of 50.47%. This decrease was maintained during the intervention period. Prior to the introduction of RAPID there was an upward trend in time to diagnosis (0.25 days/month [95% CI 0.04-0.46, p = 0.02]. There was reversal in this trend following the introduction of RAPID although the change in trend from the pre to post-intervention period was nonsignificant (-0.23 days/month [95% CI -0.52-0.06, p = 0.126].

The monthly time to diagnosis with fitted segmented linear regression lines are shown graphically in Figure 61. The graph includes the counterfactual scenario (dashed blue line) in which the model estimates the future time to diagnosis by carrying forward the slope of the pre-intervention line.

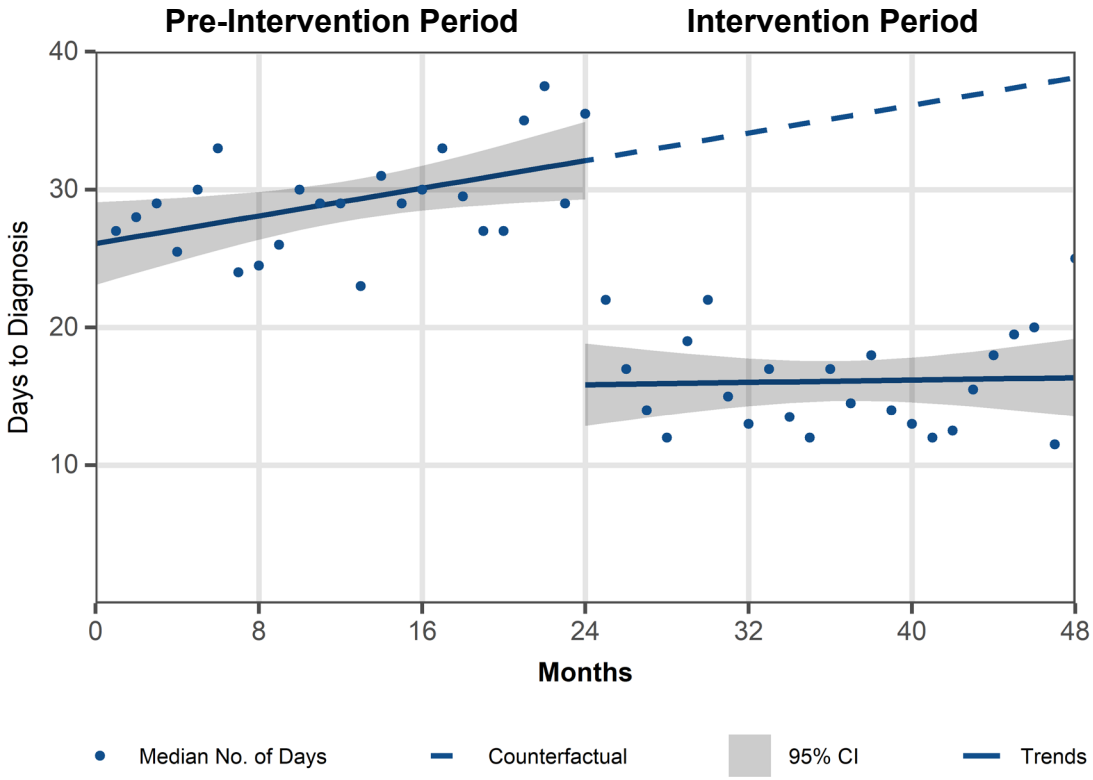


Figure 58: Interrupted time series analysis showing the median time to diagnosis before and after introduction of the RAPID pathway. The solid blue lines represent fitted estimates using a linear step change model. The counterfactual scenario is shown by the dashed blue line in which the model assumes the time to diagnosis in the absence of the RAPID pathway being introduced.

If each period is taken as an aggregate level, the median time to diagnosis during the pre-intervention period was 29.5 days (IQR 16-49) and during the intervention period was 16.0 days (IQR 8-28). The waiting time from referral to first appointment was significantly reduced but did not account for the majority of the decrease in time to diagnosis. The waiting time from referral to first appointment fell from a median 10 days (IQR 6-13) to 7 days (IQR 4-11) ($p<0.001$).

10.4.2 Site-by-site diagnostic times

A breakdown of the study population by site for the pre-intervention and intervention periods is shown in Table 49.

Table 49: Study population by site

Site	Pre-Intervention (N = 3,435)	Intervention (N = 2,130)	Total (N = 5565)
Site 1	1,084 (32%)	884 (42%)	1,968 (35%)
Site 2	1,309 (38%)	794 (37%)	2,103 (38%)
Site 3	1,042 (30%)	452 (21%)	1,494 (27%)

Statistics are presented as n (%) for categorical

Diagnostic times were consistently reduced across all sites. The time from referral to first appointment, defined as MRI or face-to-face, was reduced by 2 to 3 days depending on site. At Site 1, this reduction was from 7 (IQR 5-11) to 5 (IQR 3-7) days, Site 2 from 12 (IQR 8-14) to 9 (IQR 7-13) days and Site 3 from 9.5 (IQR 6-11) to 7 (IQR 6-10) days (Figure 62).

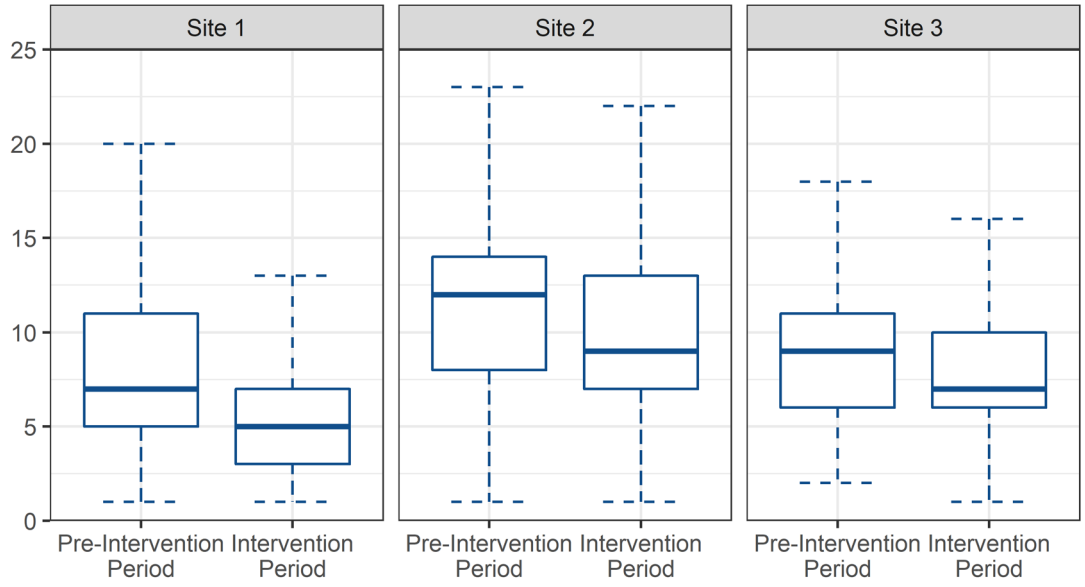


Figure 59: Time from referral to first appointment for each site in pre-intervention and intervention period. Plots show median (horizontal line), interquartile range (box) and range (whiskers). Outliers are not shown.

The median time from referral to diagnosis was significantly reduced across each site (Figure 63). The largest reduction was seen at Site 3 where the diagnosis time reduced from 30 (IQR 22-49) to 12 (IQR 7-21) days, representing at decrease of 60%. The reduction in time to diagnosis at Site 1 was 27 (IQR 16-42) to 15 (IQR 7-25) days, representing at decrease of 44%, and Site 2 was 33 (IQR 14-56) to 19 (IQR 11-40) days, representing at decrease of 42%.

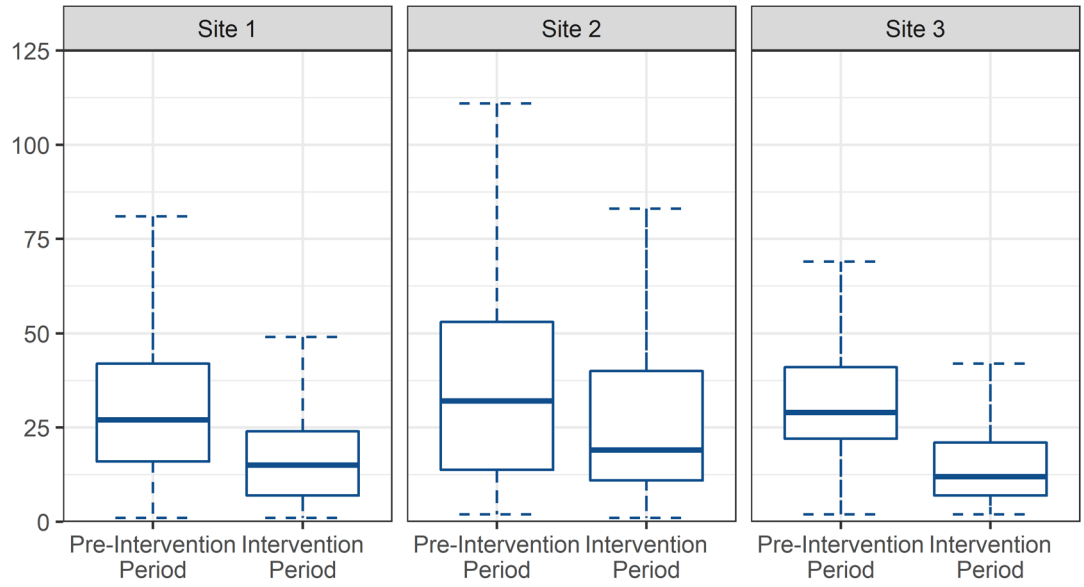


Figure 60: Time from referral to diagnosis for each site in pre-intervention and intervention period.

10.4.3 Clinical outcomes of RAPID

As secondary outcomes, the clinical outcomes of the 2,130 men who went through the RAPID pathway are reported. Table 50 shows the descriptive baseline demographic and radiological characteristics of men undergoing the RAPID pathway. The baseline age was 66yr (IQR 60-72) and PSA was 6.6ng/ml. A total of 174 (8.2%) were Afro-Caribbean ethnicity and 246 (12%) had a family history of prostate cancer. A prior biopsy for prostate cancer and taking 5α-reductase inhibitor (5ARIs) was reported in 5.1% and 3.7% of men respectively. An abnormal DRE was recorded in 287 patients (13%).

The site-by-site comparison shows significant differences in terms of age, PSA, ethnicity, family history of prostate cancer, prior prostate biopsy, 5ARIs, Abnormal DRE, MRI score, number of MRI lesions ($p<0.01$). The differences in baseline characteristics are consistent with differences in the local populations surrounding each site. With respect to age and ethnicity, participants in Site 1 were older, had a higher PSA and greater proportion of Afro-Caribbean participants than Site 3. PSA and prior biopsy for prostate cancer was highest at

baseline among men at Site 1 and a higher rate of previous family history was seen in Site 2. There were major differences across sites with respect to prostate volume (p=0.5)

Table 50: Demographic and radiological characteristics of RAPID by site

		RAPID Site		
	Overall N = 2130[1]	**Site 1** N = 884[1]	**Site 2** N = 794[1]	**Site 3** N = 452[1]
Demographic characteristics				
Age at referral (yr)	66 (60-72)	68 (61-73)	67 (60-72)	63 (58-69)
PSA (ng/ml)	6.6 (4.8-9.7)	6.6 (4.9-10.0)	6.9 (5.1-10.1)	5.8 (4.0-8.3)
Afro-Caribbean	174 (8.2%)	34 (3.8%)	94 (12%)	46 (10%)
Family history of PCa[1]	246 (12%)	129 (15%)	65 (8.2%)	52 (12%)
Prior prostate biopsy	107 (5.1%)	42 (4.8%)	52 (6.6%)	13 (2.9%)
5ARIs	79 (3.7%)	26 (2.9%)	42 (5.3%)	11 (2.4%)
Abnormal DRE	287 (13%)	165 (19%)	91 (11%)	31 (6.9%)
MRI characteristics				
Prostate Volume	58.5 (34.1)	57.9 (30.9)	59.2 (37.4)	58.6 (34.2)
MRI Score				
1-2	816 (38%)	337 (38%)	278 (35%)	201 (44%)
3 + PSAd<0.12	210 (9.9%)	121 (14%)	49 (6.2%)	40 (8.8%)
3 + PSAd>=0.12	220 (10%)	120 (14%)	55 (6.9%)	45 (10.0%)
4	440 (21%)	161 (18%)	177 (22%)	102 (23%)
5	424 (20%)	139 (16%)	224 (28%)	61 (13%)
No MRI score	20 (0.9%)	6 (0.7%)	11 (1.4%)	3 (0.7%)

Continuous data are presented as median (interquartile range) or mean (standard distribution) and categorical data as n (%)

[2] Family history defined as presence of ≥1 first- or second-degree relative with a history of prostate cancer

Abbreviations: DRE = digital rectal examination, PCa = Prostate Cancer, 5ARI = 5α-reductase inhibitor

The clinical outcomes by MRI score are described in Table 51. A non-suspicious mpMRI was reported in 48.2% (1,026 of 2130) allowing a biopsy to be omitted in 43.2% (920 of 2130). The most common MRI score reported was Score 1-2 occurring in 38.3% (816 of 2130). If the MRI score was below the suspicion threshold the biopsy rate was 10.3% (106 of 1026). Of the 106 men with a non-suspicious MRI who underwent a transperineal biopsy, 67.9% (72 of 106) were benign, 20.7% (22 of 106) were diagnosed with Gleason 3+3 and 11.3% (12 of 106) were diagnosed with Gleason ≥ 3+4. There were no cases of Gleason ≥ 4+3 in men with an MRI score 1-2 and 2 cases with MRI score 3 and PSAd <0.12.

Table 51: Clinical outcomes of RAPID pathway by MRI score

	MRI Score[1]				
	MRI Score 1-2 N = 816	**MRI Score 3 + PSAd <0.12** N = 210	**MRI Score 3 + PSAd>0.12** N = 220	**MRI Score 4** N = 440	**MRI Score 5** N = 424
Biopsy Rate	49 (6.0%)	57 (27%)	146 (66%)	386 (88%)	389 (92%)
Grade Group[2]					
Benign	33 (67%)	39 (68%)	67 (46%)	145 (38%)	55 (14%)
GG 1 (ISUP1, GS3+3)	11 (22%)	11 (19%)	26 (18%)	52 (13%)	24 (6.2%)
GG 2 (ISUP2, GS3+4)	5 (10%)	5 (8.8%)	36 (25%)	108 (28%)	102 (26%)
GG 3 (ISUP3, GS4+3)	0 (0%)	2 (3.5%)	12 (8.2%)	51 (13%)	90 (23%)
GG 4 (ISUP4, GS4+4)	0 (0%)	0 (0%)	1 (0.7%)	14 (3.6%)	37 (9.5%)
GG 5 (ISUP5, >GS8)	0 (0%)	0 (0%)	1 (0.7%)	10 (2.6%)	73 (19%)
Not reported[3]	0 (0%)	0 (0%)	3 (2.1%)	6 (1.6%)	8 (2.1%)
Outcome					
Discharged	789 (97%)	183 (87%)	136 (62%)	202 (46%)	90 (21%)
Treatment					
Active surveillance	10 (1.2%)	11 (5.2%)	38 (17%)	72 (16%)	29 (6.8%)
ADT	0 (0%)	0 (0%)	8 (3.6%)	37 (8.4%)	149 (35%)
Focal Therapy	0 (0%)	0 (0%)	6 (2.7%)	19 (4.3%)	20 (4.7%)
Other Treatment	1 (0.1%)	1 (0.5%)	5 (2.3%)	7 (1.6%)	15 (3.5%)
Radiation therapy[3]	0 (0%)	3 (1.4%)	6 (2.7%)	20 (4.5%)	50 (12%)
RRP	4 (0.5%)	2 (1.0%)	14 (6.4%)	69 (16%)	61 (14%)
Lost to follow-up	12 (1.5%)	10 (4.8%)	7 (3.2%)	14 (3.2%)	10 (2.4%)

Data are presented as n (%). Abbreviations: PSAd = PSA Density, GG = Grade Group, GS = Gleason score; , ISUP = International Society of Urological Pathology, ADT = Androgen Deprivation Therapy; RRP = Radical prostatectomy

[1] MRI stratification excludes 20 patients without an assigned MRI score for reasons such as artefact or no DCE. In this group 15 of 20 (75%) received a biopsy

[2] % calculated using men who underwent a prostate biopsy

[3] Radiation therapy includes external beam radiation and brachytherapy

The most common reasons for performing a biopsy with a non-suspicious MRI was a high PSA density in 42% (38 of 106), additional risk factors such as ethnicity or family history in 14% (13 of 106), entering a clinical trial requiring a biopsy in 10% (9 of 106), or at the patient request in 8.9% (8 of 106). The outcomes of men with a non-suspicious MRI show that the majority (94.7%) (972 of 1026) could be discharged from the pathway. There were six men who required surgical treatment with radical prostatectomy.

A suspicious mpMRI was reported in 50.2% (1,084 of 2130) which was either due to an MRI score 3 in 10.3% (220 of 2130), score 4 in 20.7% (440 of 2130) or score 5 in 19.9% (424 of 2130). If men had an MRI score above the per-protocol MRI score suspicion level, the biopsy rate was 84.8% (921 of 1084). Of the 921 men with a suspicious MRI who underwent a transperineal biopsy, 28.9% (267 of 921) were benign. The detection rate for Gleason 3+3 was 11.1% (102 of 921), Gleason ≥ 3+4 was 59.9% (552 of 921) and Gleason ≥ 4+3 was 31.4% (289 of 921). The percentage of no cancer, insignificant cancer and significant cancer found for each MRI score is shown in Figure 64.

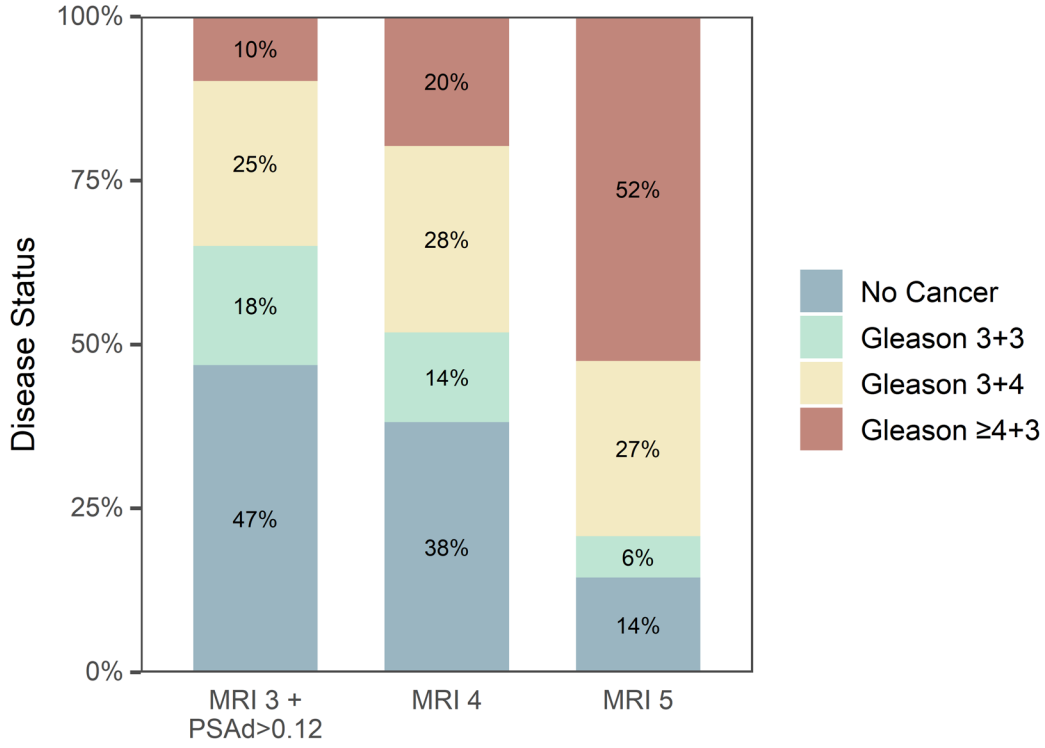


Figure 61: Percentages of men with no cancer, Gleason 3+3 and Gleason ≥ 3+4 across each to MRI score. Cumulative percentages may not equal 100% due to rounding.

10.4.4 Site variability

There was variability in clinical and diagnosis time outcomes between sites. The biopsy rate was 47.3% (418 of 884) in Site 1, 55.3% (439 of 794) in Site 2 and 40.9% (185 of 452) in Site 3. The percentage of men who were discharged from the pathway was similar at 66.1% in Site 1, 64.1% in Site 2 and 69.5% in Site 3. In men who required active treatment there was variability consistent with variability in the treatment modalities offered at each site.

Table 52: Biopsy Rate and Pathway outcomes by site

	Site 1 N = 884	**Site 2** N = 794	**Site 3** N = 452
Biopsy Rate (%)	418 (47.3%)	439 (55.3%)	185 (40.9%)
Outcome			
Discharged	584 (66.1%)	509 (64.1%)	314 (69.5%)
Treatment			
Active surveillance	65 (7.4%)	65 (8.2%)	30 (6.6%)
ADT	107 (12.1%)	87 (11.0%)	5 (1.1%)
Focal Therapy	0 (0.0%)	45 (5.7%)	1 (0.2%)
Other Treatment	13 (1.5%)	5 (0.6%)	12 (2.7%)
Radiation therapy[3]	35 (4.0%)	40 (5.0%)	8 (1.8%)
RRP	64 (7.2%)	26 (3.3%)	61 (13.5%)
Lost to follow-up	54 (2.5%)	16 (1.8%)	17 (2.1%)

Data are presented as n (%). [3] Radiation therapy includes external beam radiation and brachytherapy

Of men who had a biopsy Gleason 3+3 was detected in 11.7%, 10.3%, 16.2%, Gleason ≥ 3+4 in 60.8%, 51.3% and 43.8% and Gleason ≥ 4+3 in 35.4%, 28.5% and 14.6% in Sites 1, 2 and 3 respectively (Table 53). The rate of benign biopsy was 25.6% at Site 1, 37.8% at Site 2 and 36.8% at Site 3.

Table 53: Detection rates in men who had a biopsy by site

	Site 1 N = 418	**Site 2** N = 440	**Site 3** N = 185
Benign	107 (25.6%)	166 (37.8%)	68 (36.8%)
Gleason3+3	49 (11.7%)	45 (10.3%)	30 (16.2%)
Gleason ≥3+4	254 (60.8%)	225 (51.3%)	81 (43.8%)
Gleason ≥4+3	148 (35.4%)	125 (28.5%)	27 (14.6%)

Data are presented as n (%)

The contour plots show the positive predictive value for Gleason 3+4 according to MRI score and PSA density at each site (Figure 65). The differences in the slopes represents differences in the performance of PSAd and MRI at each site. The plots suggest that the PPV for MRI score and PSA density was similar between cohorts until the higher ranges of PSA density. The contour plots of Sites 1 and 2 are similar while Site 3 demonstrates higher risk of significant cancer in men with a PSA density above 0.3 regardless of MRI score.

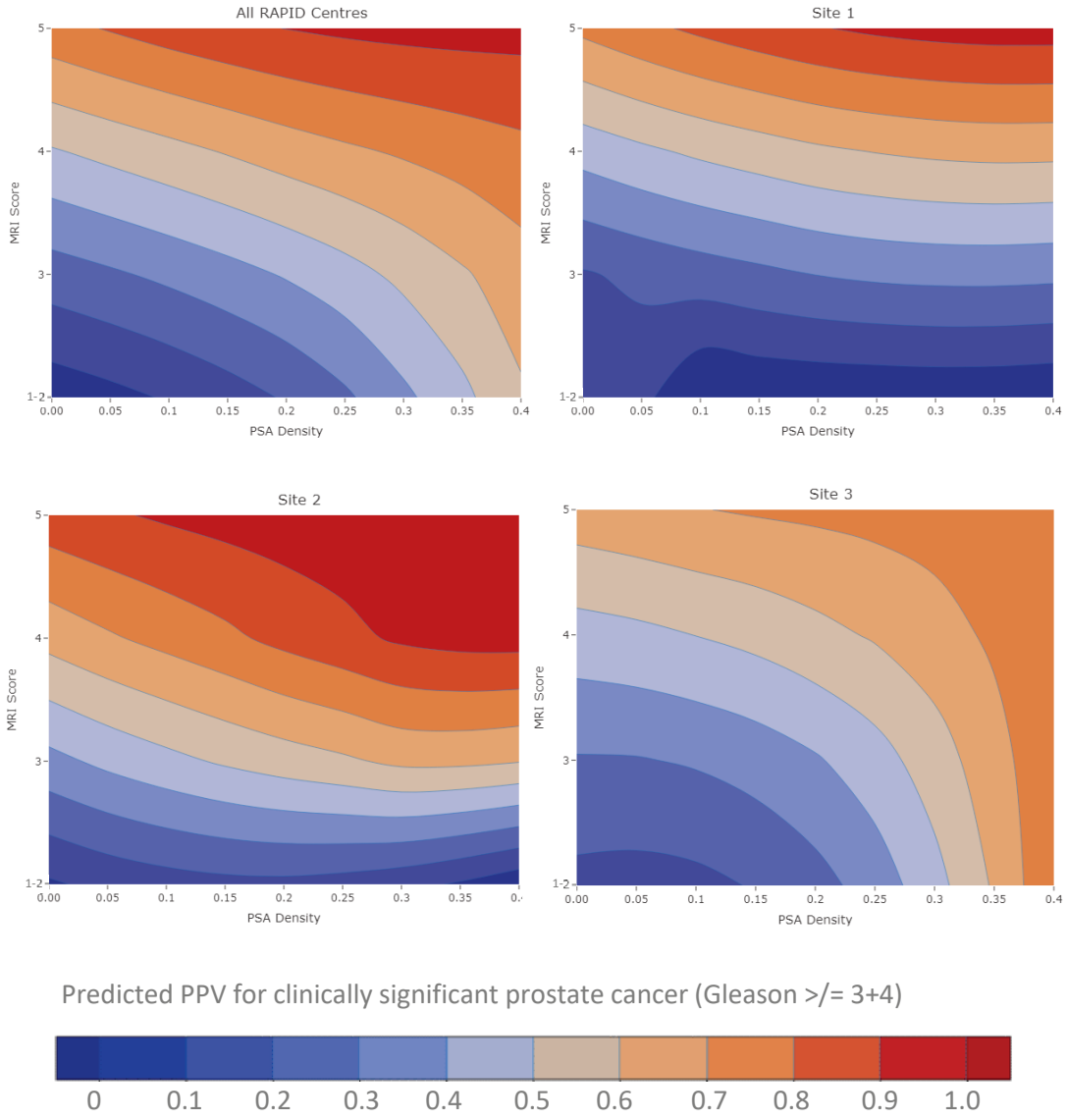


Figure 62: Contour maps for the positive predictive values obtained when varying MRI score and PSA density thresholds are applied. In each graph the x-axis is PSA density and y axis is MRI Score (PI-RADS or Likert). The positive predictive value of each combination of PSA density and MRI score is shown using the coloured map from blue (0) to red (1). For example, a PSA density of 0.2 and MRI score provides a PPV 30-40% in all RAPID centres.

Table 54 presents the variability in methods and complications of biopsy across each site. In terms of same day biopsy, only Site 2 was able to offer same-day biopsy in 24% (107 of 440). There was no availability at Sites 2 or Site 3. Each site offered different forms of analgesia for biopsy and grade of operator. Site 1 used either consultant (315 if 418) or non-consultant doctors (168 of 418) and predominately used sedation (227 of 418). Site 2 predominately delivered the biopsy pathway using non-consultant doctors (327 of 440) and was able to offer a wider range including local anaesthetic biopsy in 47% (169 of 440). Site 3 delivered biopsies using consultant doctors in 77% (104 of 185) and predominantly used general anaesthesia (88 of 185). Rates of sepsis and urinary retention were low at all sites regardless of grade of operator and technique.

Table 54: Biopsy techniques and complications per site

	Site 1 N = 418	**Site 2** N = 440	**Site 3** N = 185
Same Day Biopsy	0 (0.0%)	107 (24%)	0 (0%)
Method of analgesia			
Periprostatic local anaesthetic (LA)	45 (16%)	169 (47%)	7 (7.4%)
Intravenous Sedation	227 (78%)	153 (43%)	0 (0%)
General Anaesthetic (GA)	18 (6.2%)	35 (9.8%)	88 (93%)
Unknown	128	83	90
Biopsy Technique			
Software registration	315 (79%)	199 (53%)	90 (51%)
Cognitive registration	82 (21%)	180 (47%)	85 (49%)
Unknown	21	61	10
Biopsy Operator			
Consultant	208 (50%)	77 (18%)	104 (77%)
Non-Consultant Doctor	168 (41%)	327 (75%)	31 (23%)
Nurse Practitioner	37 (9.0%)	32 (7.3%)	0 (0%)
Unknown	5	4	50
Positive cores, median (IQR)	3 (1-5)	0 (0-4)	3 (0-6)
Total cores, median (IQR)	12 (9-14)	12 (10-15)	12 (12-12)
Complications			
Acute Urinary Retention	0 (0%)	4 (0.9%)	1 (0.5%)
Sepsis (Urine or Blood Culture Proven)	2 (0.5%)	0 (0%)	0 (0%)
Haematuria requiring irrigation	1 (0.2%)	1 (0.2%)	0 (0%)

10.4.5 Acceptability outcomes

Of 360 men who attended the RAPID pathway during the time period for collection of patient satisfaction surveys, 141 (39.2%) completed an acceptability questionnaire. The response rate was comparable across each site with Site 1, 35/120 (35.8%), Site 2 62/140 (44.3%) and Site 3 36/100 (36%). The satisfaction survey responses between sites are documented in Table 55. Overall 136 (96%) rated the RAPID experience as 'Very good or Good' and 80 (57%) reported that the time to test results were sooner than expected. After completing all tests, 70 (50%) would have wanted them on the same day and 33 (23%) wanted them on different days.

Table 55: Patient acceptability results

		RAPID Site		
	Overall N = 141	**Site 1** N = 43	**Site 2** N = 62	**Site 3** N = 36
The overall experience was				
Very Good / Good	136 (96%)	40 (93%)	60 (97%)	36 (100%)
Neither good nor poor	2 (1.4%)	2 (4.7%)	0 (0%)	0 (0%)
Fairly Poor / Very Poor	2 (1.4%)	1 (2.3%)	1 (1.6%)	0 (0%)
I don't know	1 (0.7%)	0 (0%)	1 (1.6%)	0 (0%)
The length of time from referral to MRI was				
Sooner than expected	111 (79%)	28 (65%)	50 (81%)	33 (92%)
As soon as necessary	25 (18%)	11 (26%)	11 (18%)	3 (8.3%)
Needed a bit/lot sooner	5 (3.5%)	4 (9.3%)	1 (1.6%)	0 (0%)
Too soon	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
The length of time from MRI to biopsy was				
Sooner than expected	60 (43%)	15 (35%)	36 (58%)	9 (25%)
As soon as necessary	33 (23%)	7 (16%)	20 (32%)	6 (17%)
Needed a bit/lot sooner	7 (5.0%)	5 (12%)	1 (1.6%)	1 (2.8%)
Too soon	1 (0.7%)	0 (0%)	0 (0%)	1 (2.8%)
I don't know	5 (3.5%)	1 (2.3%)	1 (1.6%)	3 (8.3%)
No biopsy	35 (25%)	15 (35%)	4 (6.5%)	16 (44%)
The length of time from referral to test results				
Sooner than expected	80 (57%)	18 (42%)	34 (55%)	28 (78%)
As soon as necessary	44 (31%)	12 (28%)	24 (39%)	8 (22%)
Needed a bit/lot sooner	13 (9.2%)	9 (21%)	4 (6.5%)	0 (0%)
I don't know	4 (2.8%)	4 (9.3%)	0 (0%)	0 (0%)
If you had to all the tests again would you prefer to have them on the				
Same day	70 (50%)	27 (63%)	17 (27%)	26 (72%)
Different days	33 (23%)	7 (16%)	25 (40%)	1 (2.8%)
I do not mind	32 (23%)	7 (16%)	20 (32%)	5 (14%)
I do not know	6 (4.3%)	2 (4.7%)	0 (0%)	4 (11%)

10.5 Discussion

10.5.1 Principle findings

This chapters reports the outcomes of the RAPID pathway after two years using a multicentre national registry. To our knowledge this chapter provides the most comprehensive analysis of the outcomes of a timed diagnostic pathway for prostate cancer which utilises a similar approach to the proposed new screening pathway.

The results show that the RAPID pathway resulted in a significant and persistent reduction in diagnostic time. The pathway reduced variations in waiting time performance between sites and it allowed 43.2% of men to avoid a prostate biopsy in contrast to the previous pathway where the majority of men would have received a TRUS biopsy. Men who received a targeted transperineal biopsy on the basis of a suspicious mpMRI had a low rate of sepsis and the detection rates for significant cancer were comparable to published series[128].

The strengths of our study include the fact that the pathway was completed by a large group of patients across three different centres. By piloting the pathway in a range of centres, comprising both secondary and tertiary units, we have shown that the RAPID pathway can be effective across a variety of hospital settings with different organisational structures and patient populations. Our findings suggest that the RAPID pathway could be a successful approach to deliver, at high-volume, prostate cancer diagnostics in a similar structure used in a potential screening programme for prostate.

This was investigated using a mixed method design which included an ITS analysis. This analytic approach is considered a strong quasi-experimental research design. ITS was performed using a segmented regression model to analyse the effect of introducing RAPID on time to diagnosis. In this model the pre-intervention period acts as the control for the post period observation and accounts for baseline levels, auto-correlation and outcome trends. The ITS analysis showed that at the point of introduction of the RAPID pathway, time to diagnosis reduced by 16.25 days which was maintained across the study period.

The analysis was performed using data from a national registry and each RAPID site agreed standardised data elements prior to commencing the pathway. The registry was implemented across the RAPID network to facilitate data collection and ensure harmonization between clinical and research workflows.

10.5.2 Comparison with previous studies

While many studies have reported the diagnostic outcomes of mpMRI[128], there have been limited previous studies evaluating the practicalities of delivering mpMRI within a timed diagnostic pathway. This lack of empirical evidence on the deliverability of the pathway creates a legitimate problem for healthcare policy makers. In this study, we have shown that it is feasible to deliver a pathway with a short diagnostic interval using mpMRI. The time series outcomes show that once the RAPID pathway is set up, faster diagnostic outcomes are likely to be sustained at two years.

Our findings are similar to a previous single centre study which piloted reserving upfront mpMRI slots to reduce diagnostic time in the prostate cancer diagnostic pathway [332]. The RAPID pathway utilised a similar concept of reserving mpMRI slots to deliver a streamlined pathway and this was augmented with same-day MRI reporting combined with, at certain sites, same-day biopsy in men with a suspicious mpMRI.

Sites which offered same-day biopsy were able to provide a one-stop pathway similar to the model utilised in one-stop diagnostic breast clinics where women are offered imaging and biopsy in a single visit[333]. This model could not be delivered at all sites due to logistical challenges and overcoming complex connections and relationships between multiple departments (urology, radiology and operating theatres). If a similar model to RAPID was adopted by new Rapid Diagnostic Centres, the processes to allow same-day biopsy could be agreed a priori and avoid the challenges associated with managing patient flows across different teams and hospital services. The increasing uptake of office-based transperineal biopsy techniques using local anaesthetic is expected to improve deliverability of same-day biopsy as more urologists are becoming experienced in outpatient transperineal biopsy[334, 335].

The number of men who avoided a biopsy was higher than in clinical trials due to the inclusion of PSA density as an additional factor in men with indeterminate mpMRI lesion (MRI Score 3). There remains uncertainty on the optimal diagnostic approach for men who have an indeterminate mpMRI lesion. A delicate trade-off exists between avoiding unnecessary biopsy, missing significant cancers and reducing overdiagnosis of insignificant cancers. In this cohort the inclusion of a PSA density threshold allowed approximately one third of men with an indeterminate lesion to avoid a prostate biopsy. The men with a low PSA density who did receive a biopsy due to other reasons had a detection rate for significant disease (Gleason ≥

3+4) of approximately 10%. This was encouraging as it is comparable to the false negative rate reported for non-suspicious MRI (Score 1-2) in diagnostic meta-analysis[336] and this rate is consistent with information provided to patients who are discharged without a biopsy regarding the risk of a false negative mpMRI. The rate of insignificant disease (Gleason 3+3) in men with indeterminate lesions was 18%, which was higher than other MRI scores highlighting the importance minimizing biopsy within this cohort.

10.5.3 Implications of findings

The results from this chapter may be useful to guide policymakers making decisions regarding the practicalities of delivering a screening pathway for prostate cancer at scale. Prior to RAPID there have been concerns that introducing MRI into the prostate cancer pathway may prolong diagnostic times. The RAPID pathway was set up in 2017 when the Department of Health in England agreed to fund a pilot at three sites across West London.

A hallmark of the RAPID pathway is the omission of prostate biopsy in the majority of men with a non-suspicious mpMRI. Standards of mpMRI reporting and follow-up were integrated into the RAPID pathway to maximise the accuracy of mpMRI and improve patient safety. The decision to omit a biopsy was re-examined within an MDT and patients could be recalled if there was clinical concern. If discharged to primary care, an individualised PSA threshold was provided to the family doctor which would warrant re-referral for repeat investigations within RAPID.

The NHS Long Term plan includes a vision of setting up Rapid Diagnostic Centres across the country and RAPID provides a one-stop model which could be delivered within this setting. This is particularly relevant given that healthcare systems are faced with the simultaneous challenge of managing increasing demand for prostate diagnostics while needing to reduce cancer waiting times. Limitations on MRI capacity can cause delays and make it difficult to deliver accelerated prostate cancer diagnostic. The demand for MRI has grown exponentially as more men are referred with a suspicion of prostate cancer. The deferral of cancer diagnostics due to COVID-19 has already increased pressure on the pathway[337].

Although a streamlined pathway is unlikely to have an impact on clinical outcomes, reducing the number of visits and tests will increase diagnostic capacity. In addition, from a patient perspective a rapid diagnosis reduces uncertainty and provides reassurance to those without the disease. This is particularly relevant in prostate cancer given the primary indication for being referred on a prostate cancer diagnostic pathway is due to a raised prostate specific

antigen (PSA) which suffers from a high false positive rate. This leads to the majority of men on the pathway being discharged without a diagnosis of prostate cancer and considerable benefit from a pathway designed to exclude cancer at the earliest point.

10.5.4 Limitations

The chapter is not without limitations. The main limitation was due to the variability between the group's baseline characteristics and the non-randomised nature of the study; any comparisons of differences in clinical outcomes were not appropriate. Although this chapter could not draw conclusions on differences in clinical outcomes, previous high-quality randomised studies have already confirmed that pre-biopsy MRI has a high sensitivity for detection of clinically significant prostate cancer. In PROMIS, the sensitivity of pre-biopsy MRI for clinically significant prostate cancer was 88% compared to 48% for TRUS biopsy[88]. This study provided level 1b evidence comparing mpMRI to TRUS using transperineal template mapping biopsies as the reference standard. PRECISION, a comparison between men only having TRUS with MRI targeted biopsy, found that significant cancer was detected in 38% in the MRI targeted biopsy group compared to 26% in the TRUS biopsy group (adjusted difference, 12 percentage points; 95% confidence interval [CI], 4 to 20; P=0.005)[133]. These findings are supported by a Cochrane meta-analysis showing that a pre-biopsy MRI pathway has superior sensitivity to the traditional TRUS biopsy pathway [336].

Second, the quasi-experimental nature of the ITS design means it is susceptible to bias. The main risk of bias in ITS analysis is that any observed change may be confounded by other improvement initiatives if they occur simultaneously with the intervention. We are not aware of any concurrent improvement initiatives during the study period.

Third, data availability prior to the implementation of the RAPID pathway meant that we were unable to compare clinical variables in the pre and post intervention periods. Following the introduction of the RAPID pathway, clinical outcomes were prospectively collected in a multicentre registry which allowed us to report the post intervention outcomes using accurate and validated data. Similar high quality data was not available prior to RAPID and would have required large-scale retrospective data collection, with data quality dependent on documentation in the patient record. Although we suspect that the majority of patients in the pre-RAPID control had a biopsy the data was not presented in this chapter.

Fourth, the patients in the pre-intervention period were primarily identified using MDT local diagnosis and discharge codes. This approach can lead to the misclassification of patients,

although a comparison between the pre and post intervention periods suggested that this bias was minimal. In addition, the eligibility criteria in the RAPID pathway may have introduced selection bias in the RAPID sample, further biasing a direct comparison of clinical outcomes.

Fifth, it is acknowledged that the study did not estimate the costs of the RAPID pathway, nor the potential savings accrued from reducing biopsy rates, sepsis and overdiagnosis. Previous studies have demonstrated the cost-effectiveness of a pre-biopsy mpMRI model[338] and further modelling work is required for a similar evaluation of RAPID. It is reasonable to think that the pathway could be cost effective given that large sections have been designed to be deliverable by nurse specialists. The role of the nurse is expanding in modern healthcare systems and our findings support results from previous studies showing that specialist nurse-led pathways can provide an effective method of delivering prostate cancer diagnostics[339].

Sixth, the study was not designed to establish the reasons for site-by-site variation in diagnostic outcomes and waiting times. Future studies could examine features which may distinguish reasons for variations in waiting times between centres. Potential key drivers include variation in pathway designs, staffing levels, equipment and infrastructure availability such as operating theatres for one-stop biopsy. Likewise, further studies could identify factors which contribute to variation in cancer detection rates such as mpMRI protocols. radiology experience, biopsy techniques and equipment availability and biopsy operator experience.

Lastly, diagnostic accuracy metrics were not calculated for each mpMRI score due to partial verification bias associated with biopsy applied in a subset of men. Further follow-up of the cohort who were discharged without a biopsy will be required to estimate diagnostic accuracy of the RAPID protocol and this will be addressed in subsequent studies with longer follow-up.

10.6 Conclusion

This chapter used a mixed method design with an interrupted time-series analysis to evaluate the outcomes of an accelerated MRI directed prostate cancer pathway. This accelerated pathway could the form the basis of a screening approach for prostate cancer. Using a multi-institutional database, the findings indicate that the RAPID pathway could be an effective model to deliver rapid prostate cancer diagnostics, similar to one-stop breast cancer clinics.

The pathway has been successfully implemented across multiple hospitals and the single visit approach was acceptable to patients. The initial clinical outcomes are satisfactory at two years pending future studies which will evaluate longer term outcomes. A limitation of delivering the pathway has been the high rate of equivocal lesions, and measures to address this warrant further investigation in the subsequent chapter.

Chapter 11 – Development of a risk model to predict significant prostate cancer in men with an mpMRI lesion

11.1 Overview

In this chapter, I describe the development of a tool to support patients making decisions regarding the need for a prostate biopsy in the context of a screen-positive MRI[x]. The performance of this tool was externally validated across multiple external, geographically distinct and independent datasets. The outcomes of this chapter have been presented at AUA[xi] (development and internal validation) and EAU[xii] (external validation results).

11.2 Introduction

In Chapter 7, it was shown that a screening prostate MRI may provide a sufficiently high sensitivity to exclude significant prostate cancer. However, the lower specificity and subsequently low PPV means that MRI can still lead to a high rate of prostate biopsy. In Chapter 6 I found that an MRI lesion may be present in 23.1% of men (Likert score ≥ 3) and up to 60% of the lesions do not have significant prostate cancer on biopsy.

Providing men who have a visible MRI lesion an estimate of their risk of harbouring cancer could provide a system to further reduce biopsy rates[340]. A personalised tool would also provide men with the opportunity to make a more individualised decision regarding the need for a prostate biopsy and allow them to more optimally balance competing risks from overdiagnosis against risk of missing significant prostate cancer.

Although MRI appears to be a sufficiently sensitive method for prostate cancer screening, the lower specificity, particularly with equivocal (MRI score 3) lesions, increases the false positive rate. When an inconclusive or equivocal MRI lesion is found, the optimal work-up is less well defined. Approximately 4 in 5 men with equivocal MRI-visible lesions do not have

[x] Acknowledgement: The model in this chapter was constructed in collaboration with Dr Max Peters and Mr Piet Kurver (University Medical Centre Utrecht). External validation was performed by Dr Ugo Falagario (University of Foggia)

[xi] Eldred-Evans, D., et al. (2020). "The Rapid Risk Score: Development of a novel risk score to predict significant prostate cancer in men with an mpMRI lesion." The Journal of Urology 203: e851-e852.

[xii] Eldred-Evans, D., et al. (2020). "The RAPID risk model: A novel risk score to predict significant prostate cancer in men with an mpMRI lesion." European Urology Open Science 19: e1741-e1742.

significant prostate cancer indicating that further upfront risk stratification could be appropriate rather than performing a biopsy in all men[257].

Numerous models have been developed which predict the risk of prostate cancer but the majority were devised prior to the development of MRI using non-targeted biopsy techniques[341]. Several contemporary models have been recently published that incorporate mpMRI findings [342-346] but some lack external validation and when it is available the benefits of the model often diminish. It is common for models to show good performance in the derivation cohort but performance often deteriorates on novel data.

Other models have been developed using specific ethnic groups and a single MRI scoring system. Given that the preferred scoring system varies between countries between either PI-RADS[180] or Likert[317], a single scoring system limits the generalisability of the models. In addition, due to the association between prostate cancer and Afro-Caribbean ethnicity, it is important that MRI models are developed considering ethnicity as a risk factor for prostate cancer.

In order to develop a contemporary and more representative model, this chapter aimed to design and externally validate a clinical prediction model for MRI lesions to be used within an MRI screening pathway.

11.3 Methods

11.3.1 Study Design

This is a prospective, multicentre cohort study conducted between 27 April 2017 and 25 October 2019 using the RAPID registry. Two university and one general hospitals participated in the development cohort for this the study. The outcomes of the registry were described in Chapter 10 and outcomes from subgroups of this cohort have been published[313, 347].

To be eligible for this study, patients in the RAPID registry had to have a pre-biopsy mpMRI with a visible MRI lesion assessed using the PI-RADS (version 2.0) or Likert scoring system. To avoid clustering bias, the highest scoring lesion was selected for subjects with multiple lesions. A visible MRI lesion was defined as PI-RADS or Likert ≥ 3. We excluded patients who did not have an MRI-directed targeted biopsy within three months of prebiopsy MRI or had a previous diagnosis of prostate cancer.

11.3.2 Predictors

A systematic search of PubMed was performed for predictors of clinically significant disease to be considered as candidate variables. A broad search was conducted for articles published on or after January 1, 2000 using the search terms ((“prostatic neoplasms”[MeSH Terms] OR (“neoplasms”[All Fields] AND “prostatic”[All Fields]) OR “prostatic neoplasms”[All Fields] OR (“prostate”[All Fields] AND “cancer”[All Fields]) OR “prostate cancer”[All Fields]) OR ((“prostate”[MeSH Terms] OR “prostate”[All Fields]) AND (“carcinoma”[MeSH Terms] OR “carcinoma”[All Fields]))) AND ((“biological markers”[MeSH Terms])) AND ((“multivariate analysis”[MeSH Terms] OR (“multivariate”[All Fields] AND “analysis”[All Fields]) OR “multivariate analysis”[All Fields]). Additional variables were considered from previous MRI derived prostate risk models published after January 1 2015[342-346].

This search identified several candidate predictor variables that are associated with risk of detection of clinically significant cancer. The predictors chosen were those that were measurable, available and reliable based on existing evidence. Four classes of variables were selected as potential predictors: demographic variables (age, Afro-Caribbean ethnicity, family history of prostate cancer), clinical variables (5α reductase inhibitors, prior negative biopsy, digital rectal examination), laboratory variables (PSA and PSA density) and radiological variables (prostate volume, PI-RADSv2 score, Likert score, combined MRI score, number of MRI lesions and index lesion size).

In terms of demographic variables; age was defined as years at the time of MRI, analysed as a continuous variable. Afro-Caribbean ethnicity was binary and considered positive in men reported as Black African, Black British, Black Other, Black Caribbean and Other Black background as per the standardised list of official ethnic groups from the Office for National Statistics in the United Kingdom. Family history was defined as any first degree relative with a diagnostic of prostate cancer.

Clinical variables included prior negative biopsy was defined as a previous transrectal or transperineal biopsy which had no cores positive for any grade of prostate cancer. Digital examination (DRE) was categorised as normal or abnormal preferentially using the examination performed in secondary care. If not performed, the DRE by the general practitioner prior to referral was reported. 5α-reductase were considered if the patient had been taking them for >6 months.

Laboratory variables included PSA was defined as the most recent level prior to prostate biopsy. PSA density was derived from PSA / Prostate volume. The number of MRI lesions included any lesion scoring Likert or PI-RADS ≥ 3.

Radiological variables were Prostate volume calculated using the exact prolate ellipsoid formula (volume = length × width × height × π/6). The MRI score of the lesion was considered as either the PI-RADSv2 or the Likert score. In cases where both scores were present, we used a composite score based on the highest scoring PI-RADSv2 or Likert lesion. In the presence of ≥2 lesions, the highest scoring lesion was selected to avoid clustering bias. The lesion size was determined by the maximal diameter of the highest score lesion. If multiple lesions of the same score were present, the largest diameter lesion was selected for analysis.

11.3.3 Data Collection and Data Quality

All participating sites prospectively collected the predictors for model development using the RAPID registry. Training was conducted prior to each site entering the registry to ensure adherence to a standardized data entry protocol. Standardised data collection forms were used with detailed definitions and instructions for each predictor variable. Quality assurance checks were performed throughout the data collection period. Any issues with data quality led to a re-review of the primary health record.

11.3.4 Reference Test

Patients underwent a targeted transperineal prostate biopsy performed using either cognitive or software registration. The procedure was performed according to a standard operating procedure used at all participating sites. When software registration was used, the BiopSee platform (Medcom,Darmstadt, Germany) was utilised. To perform software registration, MRI was imported into BiopSee® and fused with TRUS images; target lesions were contoured and biopsies taken stereotactically under real-time TRUS guidance utilising elastic registration.

Each lesion was potted individually for analysis and a minimum of three targeted cores were taken per lesion. The maximum number of cores taken was at the discretion of the operator to maximize diagnostic accuracy. Additional non-targeted systematic sampling was performed. Biopsies were performed by experienced urologists or specialist fellows. Biopsies were evaluated in accordance with the International Society of Urological Pathology standards (18) by specialist uropathologists with more than 10 years experience.

11.3.5 External Validation Cohorts

The external validation dataset include data pooled at an individual patient level from six international cohorts and five countries. These cohorts have been previously described[131, 348, 349] and all MRIs were contemporary and performed between 2013-2019. External validation was conducted in a semi-masked manner with no data pooling across cohorts 1-5 meaning that the RAPID investigators had no access to the patient-level external data. A summary of the cohorts is provided in Table 56 and detailed description of each cohort is provided in the subsequent section.

Table 56: External Validation Cohorts

No	Type of cohort	Centres, Country	MRI scanner	MRI protocol	MRI Reporting	Biopsy Technique
1	Multi-centre clinical trial	4 centres, Finland	3T (Verio/ Skyra, Siemens) or (1.5T Aera, Siemens)	Axial, sagittal T2w; 3x DWI,; axial T1w.	IMPROD bpMRI Likert*	Cognitive TRUS guided biopsies or MRI-TRUS fusion (UroNav Fusion Biopsy)
2	Multi-centre clinical trial	3 centres, Sweden and Norway	All 1.5T (Avanto-Fit/ Aera, Siemens). ly	3D volumetric interpolated T2w; axial T1w; DWI	Modified PI-RADS version 2.0#	MRI-TRUS fusion (Koelis / Artemis system (Eigen Inc.,) BioJet system (D&K Technologies GmbH,)
3	2 single centre clinical trials	1 centre, Finland	3T (Verio/ Skyra, Siemens)	Axial, sagittal T2w; 3x DWI,; axial T1w.	IMPROD bpMRI Likert*	Cognitive TRUS guided biopsies
4	Single centre consecutive series	1 centre, Milan, Italy	Various 1.5T and 3T scanners.	Axial, coronal, sagittal T2w; Axial T1w; DWI, DCE-MRI	PI-RADS version 2.0#	MRI-TRUS fusion (BioJet system, D&K Technologies GmbH,)
5	Multi-centre consecutive series	61 centres, UK	Various 1.5T and 3T scanners.	Axial, coronal, sagittal T2w; Axial T1w; DWI, DCE-MRI	Likert scoring system	MRI-Transperineal fusion (MIM-Symphony-DX, MIM Software, Cleveland, OH, USA)
6	Single centre consecutive series	1 centre, Fogia, Italy	All 1.5 T (Achieva, Philips).	Axial, coronal, sagittal T2w; axial T1w; DWI, DCE-MRI	Clinical reporting&	MRI-TRUS fusion (Navigo, UC-CARE)

11.3.5.1 Cohort 1 (MULTI-IMPROD)

This cohort is from a large, prospective, multicentre validation trial of bpMRI (MULTI-IMPROD, NCT02241122) conducted at four different institutions in Finland. These institutes were Turku, Pori, Tampere, and Helsinki. This study recruited patients between February 1, 2015, and March 31, 2017 and the primary outcomes from this study have been reported[131]. This enrolled 364 men with a clinical suspicion of prostate cancer and 154 were included in the external validation described in this chapter.

The inclusion criteria were men aged 18 years or older who had suspicion of PCa based on two repeated PSA measurements ranging from 2.5 to 20.0ng/ml and/or an abnormal DRE. Exclusion criteria were previous prostate biopsy, previous prostate surgery, previous

diagnosis of PCa, acute prostatitis, or contraindications for MRI. All men underwent a bpMRI using a body array coil. The scanner was either 3T for Turku (Verio, Siemens), Tampere (Skyra, Siemens), and Helsinki (Skyra, Siemens) or 1.5 T (Aera, Siemens) MRI in Pori.

The MRI protocol consisted of T2-weighted acquisitions in axial and sagittal planes, three DWI sequences in three separate acquisitions. No DCE was performed in this external cohort[131]. The MRIs were scored using the IMPROD bpMRI Likert scoring system which is as follows: 1) significant cancer is highly unlikely to be present; 2) significant cancer is unlikely to be present; 3) significant cancer is equivocal; 4) significant cancer is likely to be present; 5) significant cancer is highly likely to be present. The reference test was a cognitive TRUS guided biopsy or MRI-TRUS fusion. Two cores were taken from all mpMRI lesions followed by a 12-core systematic TRUS biopsy in the same session.

11.3.5.2 Cohort 2 (STHLM3 Phase 1)

This cohort consisted of patients from a prospective, multicentre, paired diagnostic study conducted at three institutions in Sweden and Norway (STHLM3-MR, NCT02788825). Men aged 45 to 75 years were eligible for inclusion if they had no prior diagnosis of prostate cancer and were referred due to a raised PSA or abnormal DRE. In this study all men had a PSA, Stockholm3 and bpMRI reported according to a modified PI-RADS score. Exclusion criteria were a prior diagnosis of prostate cancer, contraindications to MRI and severe illnesses such as metastatic cancers, severe cardio-vascular disease or dementia. This study recruited 532 men and 385 were eligible for inclusion in this study.

The MRIs were performed with a 1.5T scanner using a short version of a protocol complaint with European Society of Radiology Guidelines PI-RADS v2, in which DCE was omitted. No endorectal coils were used. MRI scans were reported according to the modified PI-RADS v2.0. Up to three lesions with PI-RADS grade ≥3 could be marked for targeted biopsies. There were six experienced uro-radiologists reviewed the MRI series. All men with PI-RADS ≥3 underwent a combined biopsy procedure with 2-3 targeted biopsies to any marked lesion, after which a systematic 12 core template biopsy was performed. Targeted biopsies were undertaken using the Koelis system (Koelis Inc.), Artemis system (Eigen Inc.), or BioJet system (D&K Technologies GmbH).

11.3.5.3 Cohort 3 (IMPROD)

This cohorts consists of pooled data from two single-centre clinical trials (IMPROD NCT01864135 and IMPROD 2.0 NCT02844829) conducted in Turku, Finland. The primary outcomes have been reported[348, 349]. IMPROD recruited between March 2013 and February 2015 and included 175 men. IMPROD 2.0 started recruitment in July 2016 and included 200 men. For our study 351 men across both studies were suitable for inclusion.

Both cohorts recruited men aged 40 to 85 years with a suspicion of prostate cancer based on a PSA from 2.5ng/ml to 20ng/ml and/or abnormal DRE. Exclusion criteria were previous biopsies in the last six months, known prostate cancer, previous prostate surgery, symptoms of prostatitis, contraindications to MRI, uncontrolled serious infection and any other conditions that might compromise patients' safety.

The MRIs were completed on a 3T scanner (Verio, Siemens, Erlangen, Germany). Sequences were the same as Cohort 1. The MRIs were reported prospectively by a single uro-radiologist with more than five years of experience according to the IMPROD bpMRI Likert scoring system. A TRUS biopsy was performed using cognitive targeting without MRI-TRUS fusion. The dominant MRI lesion was targeted with two targeted cores followed by 12 systematic TRUS biopsies.

11.3.5.4 Cohort 4 (Milan)

This cohort consists of a large prospective case series from a single tertiary care referral centre in Milan, Italy. All clinical and pathological data in the database were prospectively collected and included patients aged 45 to 74 years referred to San Raffaele Hospital, Milan between 2013 and 2017[350, 351]. The cohort included 730 consecutive patients and 570 were suitable for inclusion in this study.

MRIs were performed on either a 1.5-T mpMRI (Achieva and Achieva dStream, Philips Medical Systems) or 3-T mpMRI study (Discovery; GE Healthcare) with a phased array surface coil and endorectal coil (BPX-15; Bayer Medical Care). The imaging protocol included multiplanar T2w images, DWI with b values of 0, 800, 1400, 1600s/mm2; DCE MRI, and delayed T1-weighted images with fat suppression. The MRIs were reported according to PI-RADS v2.0 by three experienced uro-radiologists.

A software registration fusion device was used to biopsy the lesions visualized on mpMRI. This was performed using a Flex Focus 500 machine with a biplanar transducer (BK Medical, Herlev, Denmark). In the case of multiple mpMRI lesions, each was targeted followed by a standard 12-core random systematic biopsy. The systematic sampling was performed outside the mpMRI lesion at a distance of 5mm.

11.3.5.5 Cohort 5 (Fogia)

The final cohort consisted of a prospective database from a single University hospital in Foggia, Italy. The outcomes of this database are pending publication and the cohort in this study included patients who underwent a biopsy between 2015 and 2019. The database includes men with a PSA between 3ng/ml to 20.0ng/ml and/or abnormal DRE who underwent biopsy after having an mpMRI. There were 369 patients in this database and 324 were eligible for inclusion in this study.

All MRIs were performed using a 1.5 Tesla MR scanner (Achieva, Philips Healthcare) and surface array coils (SENSE Flex surface), or with an endorectal coil (ERC) combined with 16-channel surface coil (TORSO-XL coil). The MRIs were reported by two uro-radiologists with seven years' experience. All patients underwent a 12-core template systematic biopsy and in cases with a suspicious mpMRI an additional two to four targeted cores were taken guided by the Navigo MRI-US images fusion system.

11.3.5.6 Cohort 6 (MIMS)

This is a large multicentre database including MRIs from 61 centres in the United Kingdom. The outcomes of this database have been published[352]. The database includes men who underwent transperineal image-fusion targeted biopsy using MIM-Symphony-DX (MIM Software, Cleveland, OH, USA) between April 1, 2014 and June 30, 2017. The biopsies were performed at 11 centres in the UK and the indication was elevated age-adjusted prostate-specific antigen (PSA) or abnormal digital rectal examination.

A range of 1.5-T and 3.0-T scanners was used according to local MRI centre practice. All mpMRI examinations were carried out in accordance with the standards laid down by British Society of Uroradiology and the European Society of Uroradiology, with coronal or sagittal T2-weighted (T2 W) imaging of the pelvis, transverse T2 W imaging, multiple b-value apparent diffusion coefficient imaging, long b-value imaging (1500 or 2000 for 1.5 T or 3 T, respectively), and gadolinium-enhanced dynamic contrast T1-weighted axial scans. All scans

were reported using the Likert scale by experienced radiologists as set out in the UK consensus.

Patients with lesions scores 3, 4, or 5 underwent transperineal targeted and systematic biopsies; some physicians performed targeted biopsies only according to local practice. The transperineal biopsy was performed using core needles inserted via a brachytherapy grid fixed on a stepper. The protocol for the number of cores per target was four to six, although clinician discretion permitted more if appropriate. Targeted biopsies were always carried out first, with a maximum of four lesions targeted. Non-targeted biopsies involved sampling non-suspicious areas on MRI.

11.3.6 Outcomes

The primary endpoint used for the development and validation of the Rapid Risk Score was detection of any length Gleason ≥3+4. This was selected as it is a conservative definition for significant disease and its widespread availability allowed external validation. To account for the range of definitions for significant disease, multiple alternative models were constructed for secondary definitions including Gleason ≥ 4+3 and UCL/Ahmed 1 definitions of clinically significant prostate cancer.

11.3.7 Model Development

The general framework for model development consisted of transforming and selecting predictors, followed by imputation of missing data; constructing multivariable binary logistic regressions models using the backward stepwise method, and finally internally validating and evaluating model performance. The model was developed in R Version 4.0.3.

11.3.7.1 Preparation of predictors

The first step was selection of candidate predictors in their best form (categorical vs continuous) determined by graphical analysis. Normality of the continuous variables was visually assessed using QQ plots and confirmed with the Shapiro-Wilk test for normality. PSA and PSA density were transformed using natural logarithm to obtain a normal distribution and provided a better fit of the models; the other independent variables did not have to be transformed.

11.3.7.2 Missing data

To account for missing data, multiple imputations with chained equations were used in accordance with previous publications[353]. Missing data is inherent in observation cohorts and multiple imputation rather than complete case series analysis has been recommended as the preferred technique for handling missing data[354]. Multiple imputation was used, as opposed to single imputation, as it provides a number of possible datasets to control for the uncertainty produced from missing data.

The imputed datasets were then analysed using the same process as complete data and the results of each dataset were combined. In this analysis ten imputed datasets for missing variables were created using the R package MICE[355]. Missing data were considered missing at random. Variable selection was performed in all imputed datasets. All predictors and the outcomes were included in the imputation procedure. The analysis was based on twenty imputed datasets using Rubin's rules[356]. This method ensures that the uncertainty of missing data is incorporated into the model and the pooled outcomes are reliable.

11.3.7.3 Multicollinearity

The presence of linear dependencies among the candidate predictor variables was assessed using two Spearman's rho correlation coefficients. Models which include predictors that are highly correlated can lead to unstable estimates and inaccurate variances. A correlation coefficient of more than 0.7 implies the presence of multicollinearity. In cases of multi-collinearity, the least significant predictors were excluded from the model. As such, PSA density, MRI score and MRI volume were retained at the expense of PSA, PI-RADS and Likert scores respectively.

11.3.7.4 Selection of predictors and interaction terms

On the basis of the number of events (Gleason ≥ 3+4 = 683), no limitations on the number candidate predictors was required following the 1-in-10 events-per-variable rule[357]. All variables were theoretically associated with the outcomes based on clinical reasoning as well as literature. The relationship between the candidate predictor variables and the outcome of clinically significant prostate cancer was first assessed by univariable logistic regression analyses providing odds ratios (ORs) with corresponding 95% confidence intervals (CIs).

This was followed by multivariate regression to identify the final set of predictors for clinically significant cancer on biopsy using a backwards selection procedure. This approach starts with all predictors variable in the model, then sequentially drops those in order to minimize Akaike information criterion (AIC).

Interaction terms were considered between PSA / PSA density and other potential predictors in the model. Several studies have reported the association between PSA and factors such as age, Afro-Caribbean ethnicity, family history, prior prostate biopsy and prostate volume[358, 359]. The possible modifying effects of these variables to PSA were ascertained by adding interaction terms to the logistic regression models. Interactions were tested with a threshold of $p<0.05$.

11.3.7.5 Internal validation

After initial specification of each model, internal validation was performed using a standard bootstrapping procedure[360] to obtain an optimised corrected estimation of model performance. This technique provides an estimate of the degree of over-optimism of the model (ie, how much the model's performance reduces when applied to a new group of similar patients).

We re-estimated the intercept and β-coefficients in 2000 bootstrap samples from each imputation set. In each bootstrap sample the entire backward stepwise elimination modelling steps were repeated. This bootstrap procedure allowed correction for optimism and a shrinkage factor to allow correction of the final models' intercept, β-coefficients and c-statistic

11.3.7.6 Assessment of model performance

The model's performance was evaluated considering the following features:

1. Discrimination refers to the ability of the model to differentiate between a MRI visible lesion with or without significant cancer. Discrimination was quantified using the area under the curve (AUC) where values close to 1.0 indicate excellent discrimination and values close to 0.5 indicate limited discrimination ability. Differences between AUCs were compared using the DeLong method[361].

2. Calibration refers to the model's agreement between observed and predicted probability of clinically significant cancer on biopsy. It can be visualised as a

calibration plot with the observed probability on the y axis and predicted probabilities on the x-axis. A model with perfect calibration should be on the 45 degree line for agreement with the outcome. In cases of good calibration (i.e. good agreement between observed and predictive probability) all points on the calibration plot should be close to the 45 degree line. The degree of calibration was quantified using the Hosmer-Lemeshow test [362].

3. Decision curve analysis and Trade-Off Analysis was also used to assess the model's predictions on the development cohort. Decision curves are used to assess the net benefit of model's decisions at various threshold probabilities compared with the alternative of treat none/treat all clinical strategies. Trade off analysis was used to compare the reduction in number of biopsies with number of clinically significant missed.

11.3.7.7 Nomogram development

In order to increase usability of the model, a simplified scoring system was constructed using the regression coefficients of the multivariable models following the approach of Sullivan et al[363]. Using this method each predictor variable was assigned to a risk factor category and their total sum was given a predicted probability of significant cancer on biopsy

11.3.7.8 External validation

For external validation, the predicted probability for each patient in the six validation cohorts were calculated using the regression coefficients from the final models. Following similar methodology, discrimination was assessed by calculating AUC and calibration plots were constructed to assess the agreement between the predicted probabilities from the model and the observed outcome in the external cohorts. Decision curve analysis was performed to measure clinical utility.

11.4 Results

11.4.1 Study population

Between 27 April 2017 and 25 October 2019, 2,546 patients were available for model development in the RAPID registry (Figure 66). Of these 1,189 where included in the development cohort. The most common reason for exclusion was no visible MRI lesions (n = 932) and no targeted biopsy within three months (n=164). Other reasons included no PI-RADS or Likert score recorded (n = 27), a history of prostate cancer (n=15), an MRI lesion targeted scoring PI-RADS/ Likert 2 (n=17).

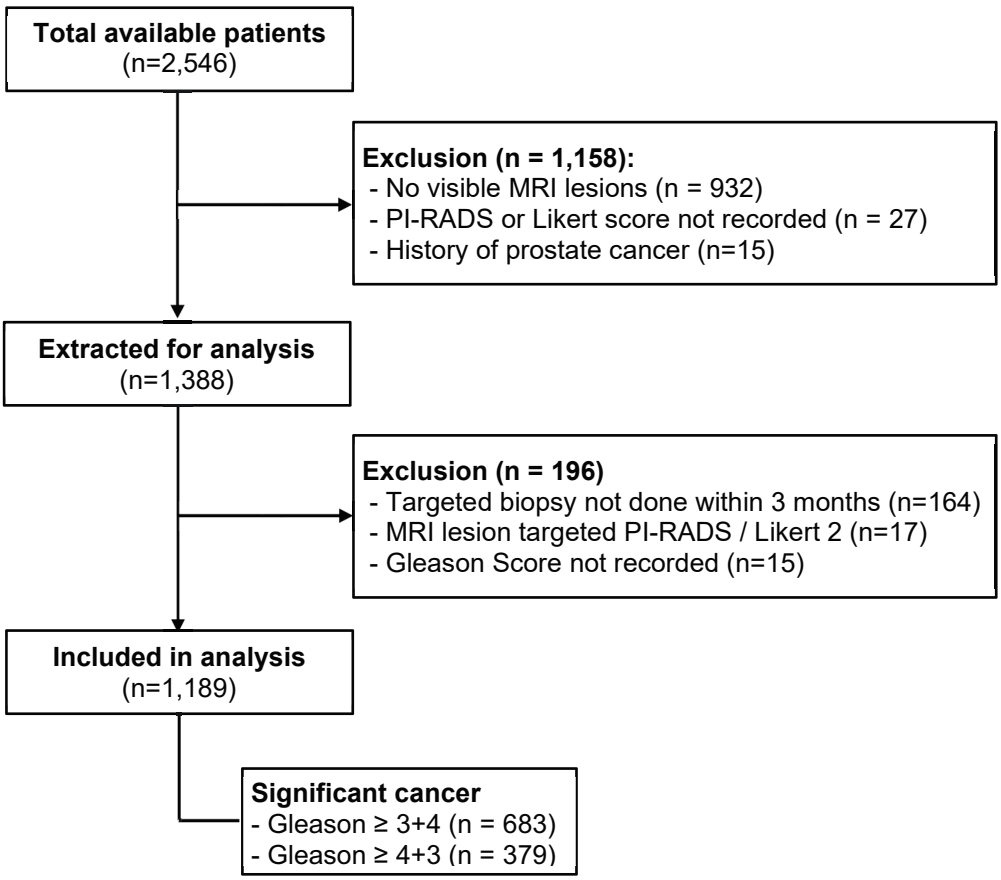


Figure 63: Flow chart for the Development Cohort

Of the total 1,189 men, 681 (57.3%) were defined as cases and 508 (42.7%) as controls on the primary outcome of Gleason ≥ 3+4. Subjects who were found to have Gleason ≥ 3+4 were 68..8 years on average (range 62-73 years) and a PSA 8.4ng/ml (range 5.9-14.0ng/ml). The descriptive and frequency statistics are summarised in Table 57 before imputation of missing data.

Table 57: Summary characteristics

Covariate	Overall (N = 1189)	Primary outcome	
		Controls (N = 508)	Cancer (N = 681)
Age (yr)	67.0 (60.5-72.0)	64.5 (59.0-69.9)	68.8 (62.1-73.4)
PSA (ng/ml)	7.3 (5.2-11.2)	6.2 (4.7-8.9)	8.4 (5.9-14.0)
PSA Density	0.2 (0.1-0.3)	0.1 (0.1-0.2)	0.2 (0.1-0.4)
Afro-Caribbean	145 (12.2%)	69 (13.6%)	76 (11.2%)
Family history of PCa	148 (12.4%)	59 (11.6%)	89 (13.1%)
Prior prostate biopsy	61 (5.1%)	42 (8.3%)	19 (2.8%)
5-ARIs	32 (2.7%)	15 (3.0%)	17 (2.5%)
Abnormal DRE	416 (42.1%)	116 (27.6%)	300 (53.0%)
Prostate Volume	42 (32-58)	49 (35-67)	38 (30-53)
Lesion Diameter	14 (10.0-21.0)	12 (9.0-18.2)	15 (10-23)
Number of Lesions			
1	835.0 (70.2%)	373.0 (73.4%)	462.0 (67.8%)
2	313.0 (26.3%)	126.0 (24.8%)	187.0 (27.5%)
3	36.0 (3.0%)	8.0 (1.6%)	28.0 (4.1%)
4	5.0 (0.4%)	1.0 (0.2%)	4.0 (0.6%)
Lesion Score			
3	251.0 (21.1%)	184.0 (36.2%)	67.0 (9.8%)
4	468.0 (39.4%)	234.0 (46.1%)	234.0 (34.4%)
5	470.0 (39.5%)	90.0 (17.7%)	380.0 (55.8%)

Abbreviations: Digital Rectal Examination (DRE), 5-Alpha Reductase Inhibitors (5-ARIs)
Summary statistics are presented as median (25%-75%) for continuous data and n (%) for categorical data

Covariate data was obtained for all 1,189 men included in the analysis. Initial examination revealed that the distribution of PSA and PSA density were not sufficiently symmetric for the normality assumptions of the model to be met. The Shapiro-Wilk normality test for PSA ($W = 0.16358$, p-value < 0.001 and for PSA density $W = 0.14834$, p-value < 0.001); all other continuous covariates were normally distributed as per Shapiro-Wilk. Due to the left skewed distribution, PSA and PSA density were logged transformed (natural log) to obtain a normal distribution.

Following the log transformation, continuous covariates were initially assessed by box and whisker plots against the domain of detection of Gleason ≥ 3+4 (primary outcome) and Gleason ≥ 4+3 (secondary outcome)

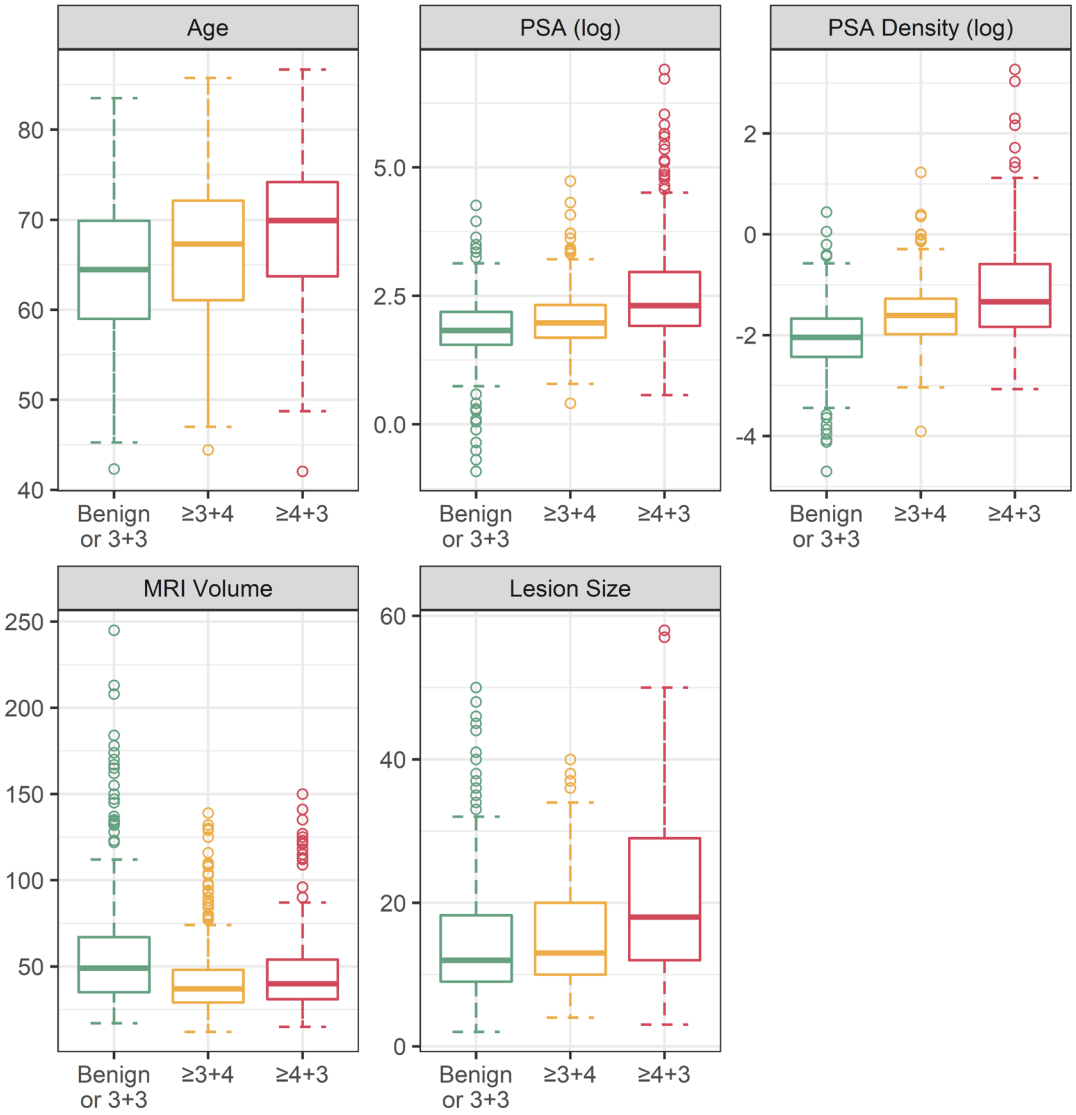


Figure 64: Feature plot for continuous co-variates using box and whisker plots stratified by Benign/Gleason 3+3, Gleason ≥ 3+4 (primary outcome) and Gleason ≥ 4+3 (secondary outcome).

Medical chart review was able to limit the percentage of missing data in the majority of covariates. The covariates which had residual missing data were DRE (16.9%), Likert Score (9.5%) and PI-RADS (0.42%). The missing data for these predictors were imputed using multiple imputations, based on all predictors and outcome status.

11.4.2 Multicollinearity

Following imputation, we explored the relationship between clinically significant cancer and the 13 predictor variables in the dataset. The pairwise correlation between potential variables was shown using a correlation matrix (Figure 68) where a higher level of correlation is represented by a larger box. Collinearity was evaluated by Spearman's rho correlation coefficients.

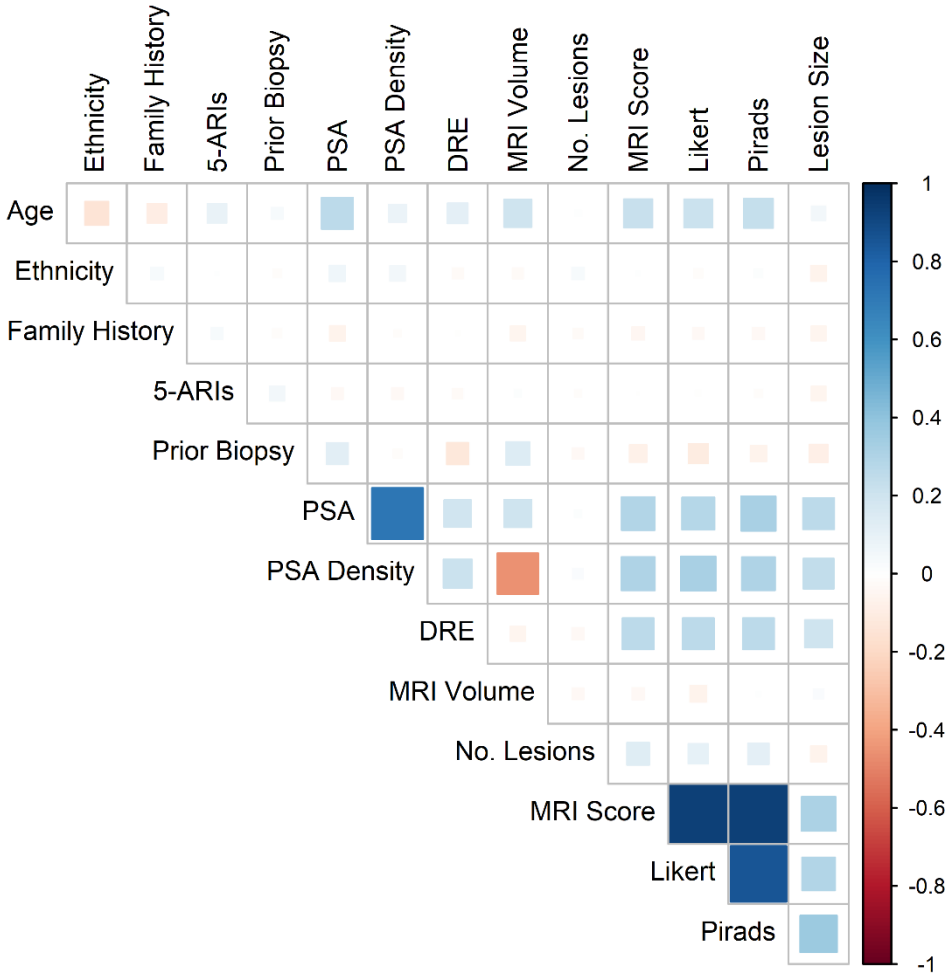


Figure 65: Correlation matrix showing the Spearman's correlation among potential predictors. Variables positively correlated appear in blue, while those negatively correlated appear in red. A threshold of r = +/- 0.7 was applied to exclude predictors which showed evidence of multicollinearity.

The correlation matrix indicated collinearity between the following variables PSA and PSA density (r=0.75), MRI score and Likert (r = 0.95), MRI score and PI-RADS (r = 0.94), Likert and PI-RADS (r = 0.88). As such, PSA density and MRI score were retained at the expense of PSA, PI-RADS and Likert respectively. MRI score was chosen as the most clinically relevant variable as it is a composite variable which incorporated by both scoring system. PSA density and MRI volume had a coefficient of 0.47 so both variables were retained in the model.

11.4.3 Evaluation of covariates and interaction terms

Univariate analysis: Each remaining covariate was assessed by univariate regression against the outcome measures in order to inform which co-variates could enter the multivariate regression. Detection of Gleason ≥3+4 was significantly associated with age (OR = 1.06, p<0.001), logged PSA density (OR = 4.11, p<0.001), prior prostate biopsy (OR = 0.32, p<0.001), prostate volume (OR = 0.98, p<0.001), Lesion size (OR = 1.03, p<0.001), number of lesions (OR 1.35, p<0.001) , MRI score 4 compared to MRI score 3 (OR = 2.75, p<0.001) and MRI score 5 compared to MRI score 3 (OR = 11.76, p<0.001).

Table 58: Univariate Regression for Gleason ≥3+4 & Gleason ≥4+3

Covariate	Primary: Gleason ≥ 3+4		Secondary: Gleason ≥ 4+3	
	OR (95% CI)	p value	OR (95% CI)	p value
Age	1.06 (1.04 - 1.08)	<0.001	1.06 (1.05 - 1.08)	<0.001
PSA Density (log)	4.11 (3.33 - 5.1)	<0.001	3.35 (2.77 - 4.10)	<0.001
Afro-Caribbean	0.80 (0.56 - 1.13)	0.2	1.00 (0.68 - 1.44)	>0.9
Family history of PCa	1.14 (0.81 - 1.63)	0.5	1.11 (0.77 - 1.59)	0.6
Prior prostate biopsy	0.32 (0.18 - 0.55)	<0.001	0.36 (0.16 - 0.70)	0.005
5-ARIs	0.84 (0.42 - 1.72)	0.6	0.84 (0.36 - 1.76)	0.7
Abnormal DRE	2.97 (2.27-3.89)	<0.001	3.34 (2.54-4.41)	<0.001
Prostate Volume	0.98 (0.98 - 0.99)	<0.001	0.99 (0.99 - 1.00)	0.003
Lesion size	1.03 (1.02 - 1.05)	<0.001	1.05 (1.04 - 1.07)	<0.001
Number of Lesions	1.35 (1.09 - 1.67)	0.007	1.13 (0.91 - 1.40)	0.3
MRI Score				
3	Ref	Ref	Ref	Ref
4	2.75 (1.98 - 3.85)	<0.001	3.79 (2.27 - 6.71)	<0.001
5	11.6 (8.12 - 16.8)	<0.001	17.0 (10.4 - 29.8)	<0.001

Abbreviations: Digital Rectal Examination (DRE), 5-Alpha Reductase Inhibitors (5-ARIs), Odds Ratio (OR), PCa = Prostate Cancer

The detection of Gleason ≥4+3 was significant related to Age (OR = 1.06, p<0.001), log PSA density (OR = 3.35, p<0.001), prior prostate biopsy (OR = 0.36, p=0.005), prostate volume (OR = 0.99, p=0.003), lesion size (OR = 1.05, p<0.001), number of lesions MRI score 4 compared

to MRI score 3 (OR = 3.79, p<0.001) and MRI score 5 compared to MRI score 3 (OR = 17.00, p<0.001). A similar univariate analysis for UCL/Ahmed 1 definition of clinically significant cancer is shown in Table 59.

Table 59: Univariate Regression for UCL/Ahmed 1

Covariate	OR (95% CI)	p value
Age	1.05 (1.04 - 1.07)	<0.001
PSA Density (log)	4.26 (3.45 - 5.33)	<0.001
Afro-Caribbean	0.87 (0.61 - 1.23)	0.4
Family history of PCa	1.20 (0.85 - 1.71)	0.3
Prior prostate biopsy	0.38 (0.21 - 0.65)	<0.001
5-ARIs	0.45 (0.21 - 0.92)	0.033
Abnormal DRE	2.51 (1.98 - 3.20)	<0.001
Prostate Volume	0.98 (0.98 - 0.99)	<0.001
Lesion size	1.04 (1.03 - 1.05)	<0.001
Number of Lesions	1.22 (0.99 - 1.51)	0.067
MRI Score		
3	Ref	Ref
4	2.57 (1.83 - 3.65)	<0.001
5	12.8 (8.91 - 18.6)	<0.001

Abbreviations: Digital Rectal Examination (DRE), 5-Alpha Reductase Inhibitors (5-ARIs), Odds Ratio (OR)

Interaction Terms: When assessing for interaction terms we hypothesized that there could be an association between Afro-Carribean ethnicity and other potential predictors in the model. Interactions between Afro-Caribbean ethnicity with all other factors were tested in an interaction analysis. Interactions were tested with a threshold of p<0.05. A significant interaction term was identified between Afro-Caribbean ethnicity and age and therefore was included within each model.

11.4.4 Model Development and Internal Validation

Multivariate modelling was conducted via a backward selection procedure which included all covariates a priori, and the optimal model was selected to minimize the AIC. Three models were developed for each definition of clinically significant cancer. Model 1 (simple) selected

co-variates from specific groups of predictors. If multiple co-variates were significant in a group, the most clinically relevant variable was selected. Model 2 (non-invasive) excluded digital rectal examination a priori as it is an invasive test which requires an additional hospital visit and has high interobserver variability. The remaining co-variates in the non-invasive model could be completed remotely or in the radiology department. Model 3 (full) included all co-variates and only excluded variables based on the backward selection AIC criterion. A summary of each model's performance is in Table 60 and the full coefficients for each model are provided in Appendix IV.

Table 60: Summary performance of each model

	AIC	AUC (95% CI)
Gleason ≥ 3+4		
Model 1 (simple)	1218.4	0.823 (0.799-0.846)
Model 2 (non invasive)	1208.2	0.829 (0.819-0.852)
Model 3 (Full)	1195.4	0.834 (0.812-0.857)
Gleason ≥ 3+4		
Model 1 (simple)	1144.9	0.814 (0.789-0.839)
Model 2 (non invasive)	1136.2	0.819 (0.793-0.844)
Model 3 (Full)	1124.6	0.823 (0.797-0.8482)
UCL/Ahmed 1		
Model 1 (simple)	1224.3	0.826 (0.803-0.849)
Model 2 (non invasive)	1216.5	0.829 (0.807-0.853)
Model 3 (Full)	1211.8	0.831 (0.809-0.854)

Abbreviations: Akaike's information criteria (AIC), Area under the curve (AUC),
Notes: A low value for AIC indicates a close fi t of the model to the true odds. A high AUC indicates better discriminator power.

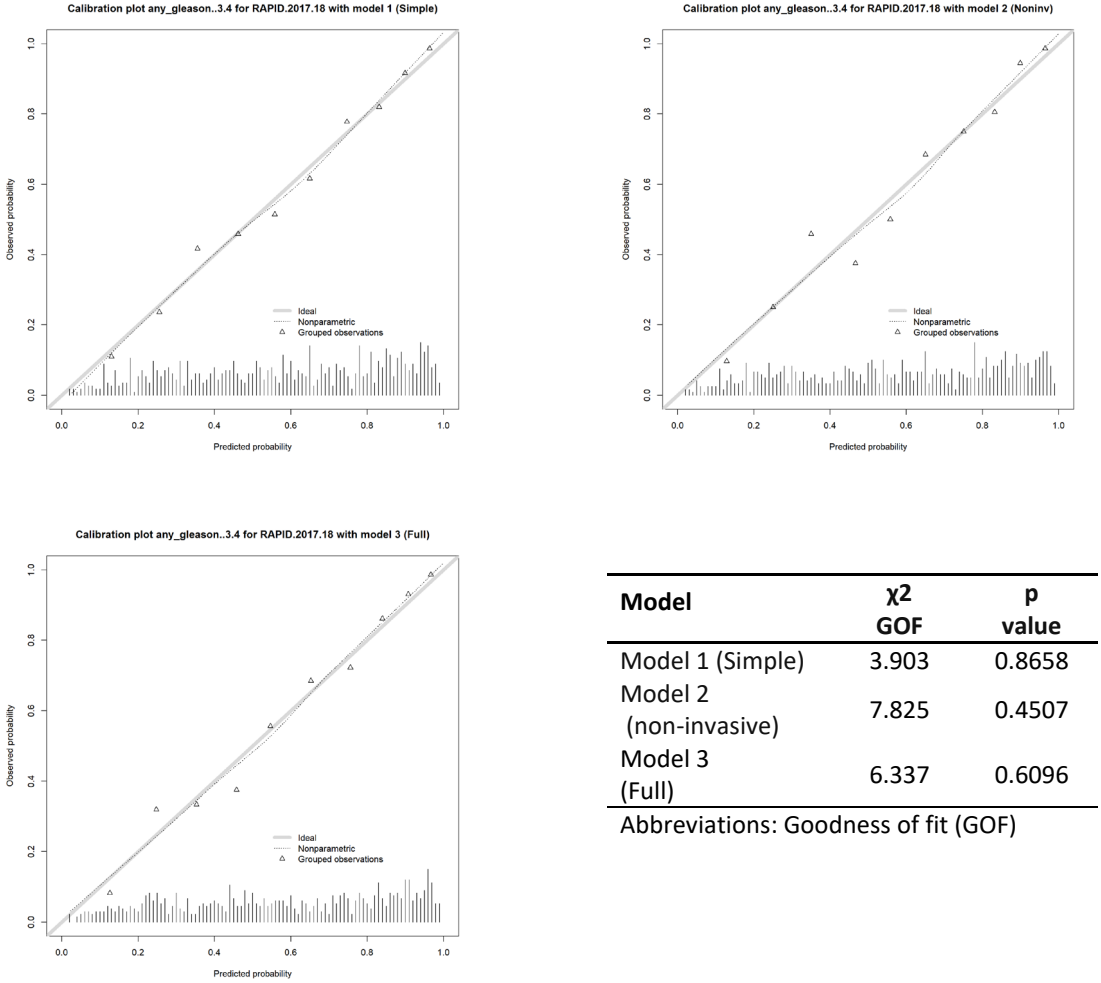
11.4.4.1 Gleason ≥3+4 (primary outcome)

The selection procedure led to the exclusion of number of lesions (p = 0.43), lesion size (p = 0.59), 5-ARIs (p = 0.45), Afro-Caribbean ethnicity (p=0.355) and abnormal DRE. The following co-variates were retained in model 1; age (β = 0.052, p=0.004), PSA density (β = 1.106, p<0.001), prior biopsy (β = -1.023, p = 0.02), MRI volume (β = -0.008, p=0.01), MRI score 4 (β = 0.935, p<0.001) and MRI score 5 (β = 1.891, p<0.001).

For model 2 (non-invasive) back variable selection identified the age (β = 0.065, p<0.001), PSA Density (log) (β = 1.22, p<0.001), prior biopsy (β = -1.05, p=0.002), MRI volume (β = -0.008, p=0.015), MRI score 4 (β = 0.0953, p<0.001), MRI score 5 (β = 1.91, p<0.001), Afro-Carribean ethinicity (β = 4.2, p=0.008) and family history (β = 0.386, p=0.003) as predictors for Gleason ≥ 3+4. Abnormal digital rectal examination was excluded a priori.

In Model 3 (full), the backward selection identified 9 variables and one interaction term associated with Gleason ≥ 3+4; age (β = 0.064, p<0.001), PSA Density (log) (β = 1.09, p<0.001), prior biopsy (β = -0.91, p=0.007), MRI volume (β = -0.008, p=0.012), MRI score 4 (β = 0.0944, p<0.001), MRI score 5 (β = 1.81, p<0.001), Afro-Caribbean ethnicity (β = 4.06, p=0.01) and family history (β = 0.388, p=0.07) and abnormal digital rectal examination (β = 0.561, p<0.001).

Visual inspection of the calibration plots for each showed reasonable agreement across the entire range of predicted risks for Gleason ≥ 3+4 (Figure 69). To compare the calibration of the model, the Hosmer-Lemeshow (HL) test was used. Model 1 predicted significant cancer with an HL test 3.903, p = 0.87. Model 2 had an HL test of 7.825, p = 0.45. Model 3 had HL test of 6.3368, p = 0.6096). The higher p values for Model 1 indicated that this model had the best fit.



Model	χ2 GOF	p value
Model 1 (Simple)	3.903	0.8658
Model 2 (non-invasive)	7.825	0.4507
Model 3 (Full)	6.337	0.6096

Abbreviations: Goodness of fit (GOF)

Figure 66: Calibration plots for Model 1-3. Calibration plots showing the relationship between the model event probability for detection of Gleason ≥ 3+4 versus average observed probability for each decile of risk. All deciles had >10 events observed per group. These curves include recalibration of the baseline models following external validation with the MIMS data. The curves indicate that Model 1 has the best fit with observed probability. This is supported by the Hosmer-Lemeshow Chi square test where Model 1 had the highest p-value

The clinical utility of each model was assessed in terms of decision curve analysis and trade off analysis. For the decision curve analysis, all the models displayed consistent positive net benefits for risk thresholds above 10% when compared with a 'biopsy in all' vs. 'biopsy in none' approach. The clinically utility of the decision curve analysis is more pronounced at higher risk thresholds. The trade off analysis showed that at thresholds above 10%, there was potential to reduce the number of biopsies by 20% but at a trade off of 1 in 10 Gleason ≥ 3+4

being missed. Given that there was no superiority for any model by clinically utility, Model 1 (simple model) was selected for nomogram development and external validation.

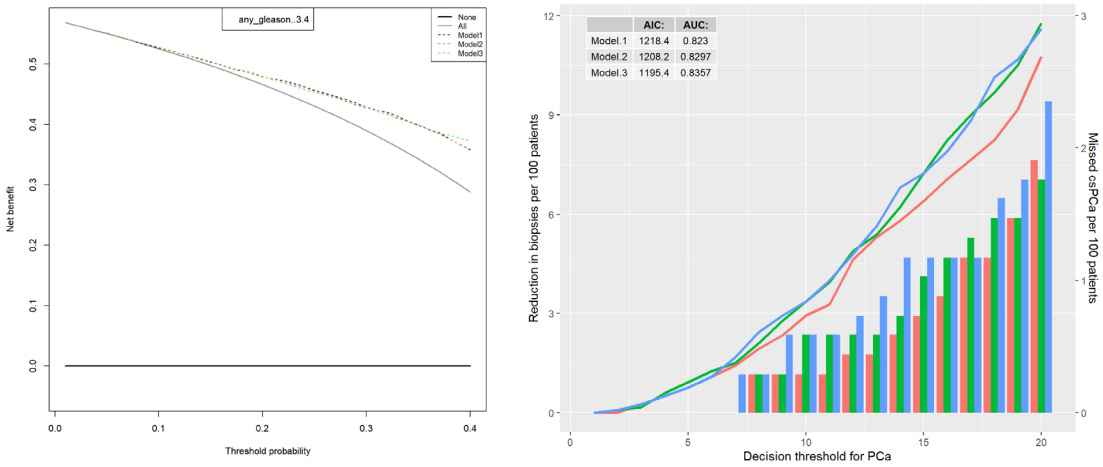


Figure 67: Decision curve analysis and Trade-Off analysis of each model. A: Decision Curve Analysis where net benefit is defined as the sum of TP (true positives) – FP (false positives) weighted by each risk threshold. For each risk threshold a larger net benefit indicates a high number TP without increasing the FP rate. Not using the model is illustrated by the alternatives of biopsy-in-all and biopsy-in none.

11.4.4.2 Gleason ≥4+3

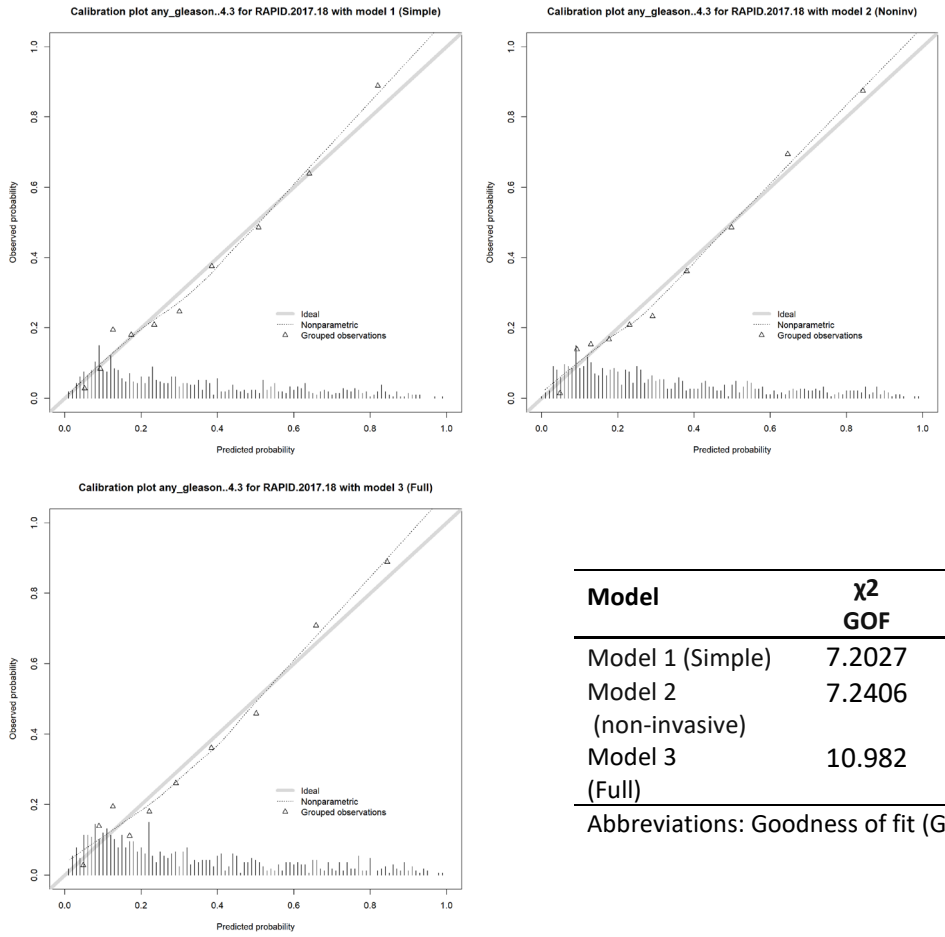
For model 1 (simple), the following variables were retained in the model age (β = 0.052, p=0.004), PSA density (β = 1.106, p<0.001), prior biopsy (β = -1.023, p = 0.02). MRI volume (β = -0.008, p=0.01), MRI score 4 (β = 0.935, p<0.001) and MRI score 5 (β = 1.891, p<0.001). The variables which were excluded included number of lesions (p = 0.43), lesion size (p = 0.59), 5-ARIs (p = 0.45), Afro-Caribbean ethnicity (p=0.355) and abnormal DRE. This model had an AIC value of 1144.9.

The multivariate regression analysis for Model 2 (non-invasive) was found to have an AIC value of 1136.2. Age (β = 0.052, p<0.001), PSA Density (log) (β = 1.13, p<0.001), prior biopsy (β = -1.03, p=0.002), MRI volume (β = -0.007, p=0.020), MRI score 4 (β = 0.0961, p<0.001), MRI score 5 (β = 1.92, p<0.001), Afro-Carribean ethinicity (β = -0.388, p=0.07) and family history (β = 0.395, p=0.07) as predictors for Gleason ≥ 3+4. Abnormal digital rectal examination was excluded a priori.

In Model 3 (full), the backward selection identified 9 variables associated with Gleason ≥ 3+4; age (β = 0.052, p<0.001), PSA Density (log) (β = 1.13, p<0.001), prior biopsy (β = -1.03, p=0.002), MRI volume (β = -0.007, p=0.020), MRI score 4 (β = 0.0961, p<0.001), MRI score 5 (β = 1.92, p<0.001), Afro-Caribbean ethnicity (β = -0.388, p=0.07) and family history (β =

0.395, p=0.07) and abnormal digital rectal examination (β = 0.571, p<001). This model had the lowest AIC value 1124.6.

When the models were calibrated to the development dataset all showed good calibration for the detection of Gleason ≥ 4+3. For Model 1 the HL test yields a non-significant p-value (7.2027, p=0.5149). Model 2 had a HL test of 7.2406, p=0.5109 and Model 3 had a HL test of 10.982, p=0.2027.



Model	χ2 GOF	p value
Model 1 (Simple)	7.2027	0.5149
Model 2 (non-invasive)	7.2406	0.5109
Model 3 (Full)	10.982	0.2027

Abbreviations: Goodness of fit (GOF)

Figure 68: Calibration plots of each clinical prediction model for Gleason ≥ 4+3. The diagonal line shows the expected versus perfect observed slope of 1 suggesting that Model has the highest level of calibration

Decision curve analysis and trade-off analysis was performed to determine the clinical usefulness of each model. The DCA analysis offered a net benefit over the 'biopsy-in-all' strategy at a threshold probability >0.05% which is higher than the benefit threshold for the previous definition of Gleason ≥ 3+4. For example, with a threshold probability of 20%, the risk models provided an added net benefit of 0.134 compared to the biopsy-in-all strategy. For the trade-off analysis a risk threshold of 20% led to a 40% reduction in biopsies but at the

expense of missing 15% of Gleason ≥ 4+3. The DCA and trade-off analysis were similar for all models so model 1 was selected as the optimum.

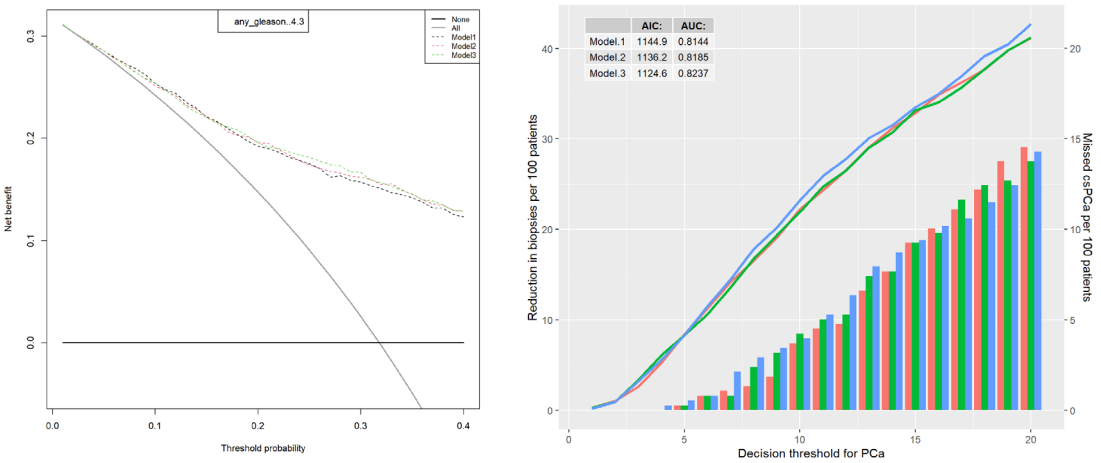


Figure 69: Models comparison by Decision Curve analysis and Trade off Analysis (a) DCA analysis shows the net benefit versus the threshold probability. The decision curve demonstrates that if the threshold probability is >5%, using the risk model for prediction of significant lesions adds more benefit than biopsying all or no patients. (b) Trade-Off shows the trade off from avoiding a biopsy at different threshold.

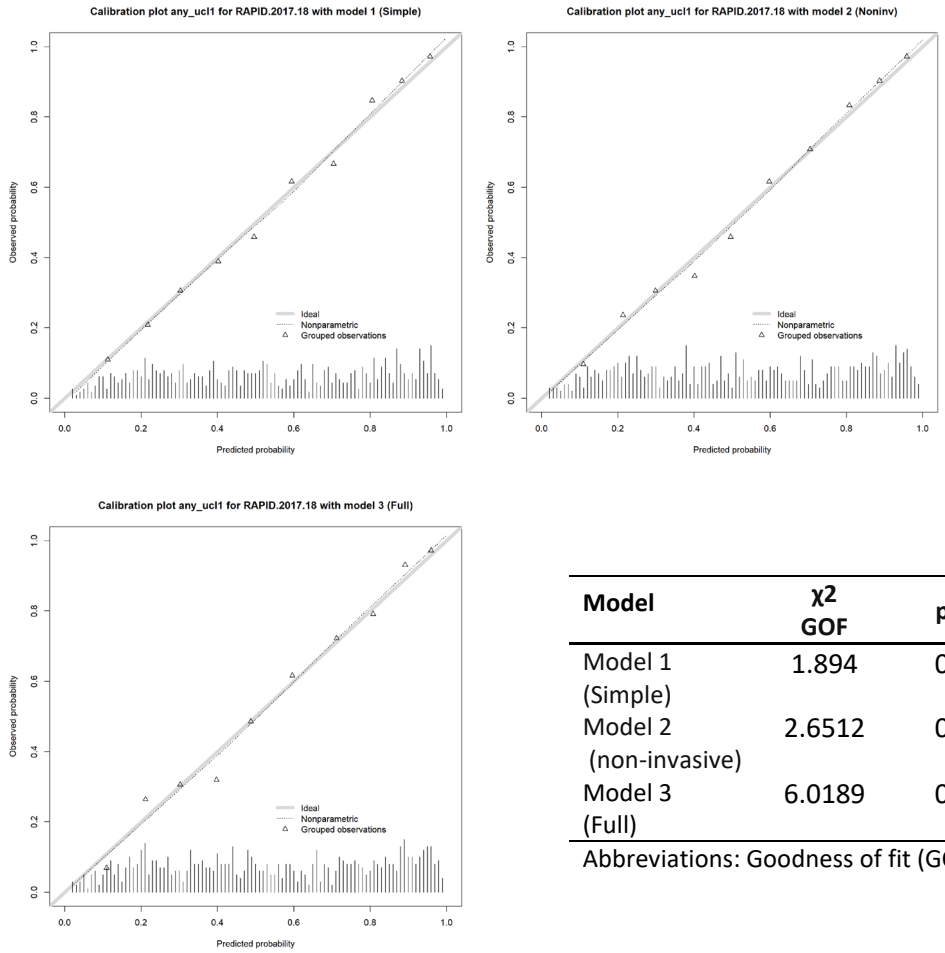
11.4.4.3 UCL/Ahmed 1

In Model 1 (simple), the variable selection process identified six variables associated with UCL/Ahmed 1 definition for clinically significant cancer; age (β = 0.041, p<0.01), PSA density (β = 1.117, p<0.001), prior biopsy (β = -0.772, p = 0.02). MRI volume (β = -0.008, p=0.02), MRI score 4 (β = 0.868, p<0.001) and MRI score 5 (β = 2.019, p<0.001).The following variables were excluded from the model; number of lesions (p = 0.43), lesion size (p = 0.59), 5-ARIs (p = 0.45), Afro-Caribbean ethnicity (p=0.355) and abnormal DRE. The AIC value was 1224.3.

For model 2 (non-invasive) variable selection identified the Age (β = 0.046, p<0.001), PSA Density (log) (β = 1.11, p<0.001), prior biopsy (β = -0.702, p=0.038), MRI volume (β = -0.007, p=0.18), MRI score 4 (β = 0.861, p<0.001), MRI score 5 (β = 2,047, p<0.001), 5-ARIs (β = -1.08, p=0.16) and family history (β = 0.478, p=0.025) as predictors for Gleason ≥ 3+4. Abnormal digital rectal examination was excluded a priori. The model had an AIC value of 1136.2.

Model 3 included nine variables; age (β = 0.045, p<0.001), PSA Density (log) (β = 1.085, p<0.001), prior biopsy (β = -0.606, p=0.074), MRI volume (β = -0.008, p=0.015), MRI score 4 (β = 0.854, p<0.001), MRI score 5 (β = 1.98, p<0.001), 5-ARIs (β = -1.109, p=0.013) and family history (β = 0.479, p=0.026) and abnormal digital rectal examination (β = 0371, p<0.014). This model had the lowest AIC value of 1124.6.

The prediction models had good calibration and useful discrimination. All calibration curves were closer to the 45 degree line and the HL test suggested that each model fitted well across different deciles. Model 1 appeared to demonstrate the highest levels of discrimination and this was confirmed by the HL goodness of fit test. For model 1 the AUC was 0.826 (95% CI 0.803-0.849 and HL test was 1.894, p=0.9841. Model 2 had an AUC of 0.829 (95% CI 0.807-0.853) and HL test of 2.6512, p=0.9543 and Model 3 had an AUC of 0.829 (95% CI 0.807-0.853).



Model	χ^2 GOF	p value
Model 1 (Simple)	1.894	0.9841
Model 2 (non-invasive)	2.6512	0.9543
Model 3 (Full)	6.0189	0.6451

Abbreviations: Goodness of fit (GOF)

Figure 70: Calibration plots for predicting UCL ≥ 1 for each model.

Similar to previous definitions of clinically significant disease, DCA and trade off analysis suggested that Model 1 (simple) would provide similar clinical utility to differentiate benign and malignant MRI lesions. The DCA analysis for the three models did not show any difference in net benefit across a range of threshold probabilities. Similarly, the trade-off analysis was consistent for the majority of risk threshold with a suggestion that Model 1 might miss more significant disease at higher thresholds

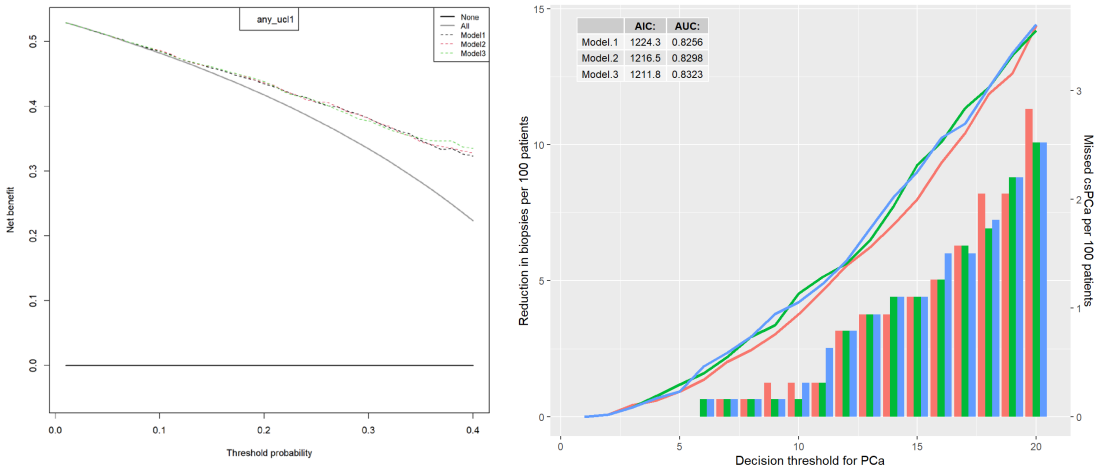


Figure 71: Decision curve analysis and Trade off analysis for the three models in the UCL 1 development cohort. (A) Shows the net benefit of using either model as a decision strategy. The line that has the highest over the widest range indicates the model with the greatest net benefit. For UCL 1 there was little difference in the net benefit between all models. (B) Trade-Off analysis for the reduction in biopsies and missed disease

11.4.5 Nomogram development

Based on the consistent outcomes of the DCA and trade-off analysis, Model 1 (Simple) was selected for nomogram and external validation. The inverse logistic function was used to estimate risk of significant cancer based on all predictor variables. The final model had the following equation for Gleason ≥ 3+4:

$$p = \frac{1}{1 + \exp(-(\alpha + \sum_i \beta_i x_i))}$$

where xi are predictor variables, β_i are the regression coefficients (see Appendix IV) and p is the probability of significant disease. The nomogram was derived from the simple model using this equation. A nomogram is a visual representation of the regression model, rather than an equation it uses a point-based system. To use the nomogram entails drawing a line from each variable up to the points scale, summing each point and circling the value on 'risk of significant cancer scale'. The final points scale will vary for each definition of clinically significant prostate cancer. Other definitions of significant cancer have similar nomograms.

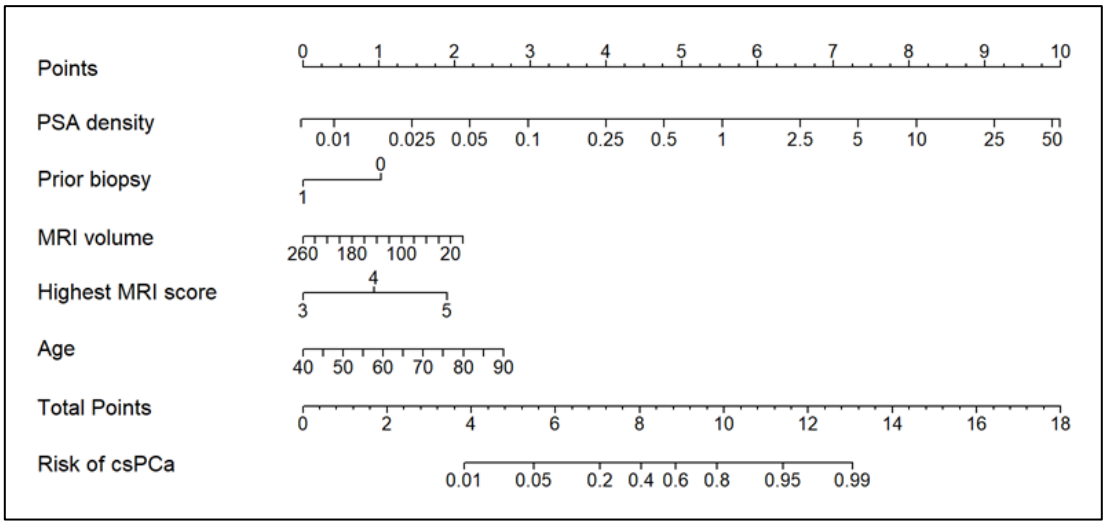


Figure 72: Nomogram for predicting clinically significant prostate cancer in men with an MRI lesion. For example, a 60 year old man would score one point and the procedure would then be repeated for each variable in the nomogram. After all the point scores are calculated, sum the total points and draw a straight line down from 'points' scale to determine the patients predicted probability of clinically significant prostate cancer.

To facilitate easy access and calculation, a web version of this nomogram was created using the Shiny application and an app was developed using the R language with the Shiny package.

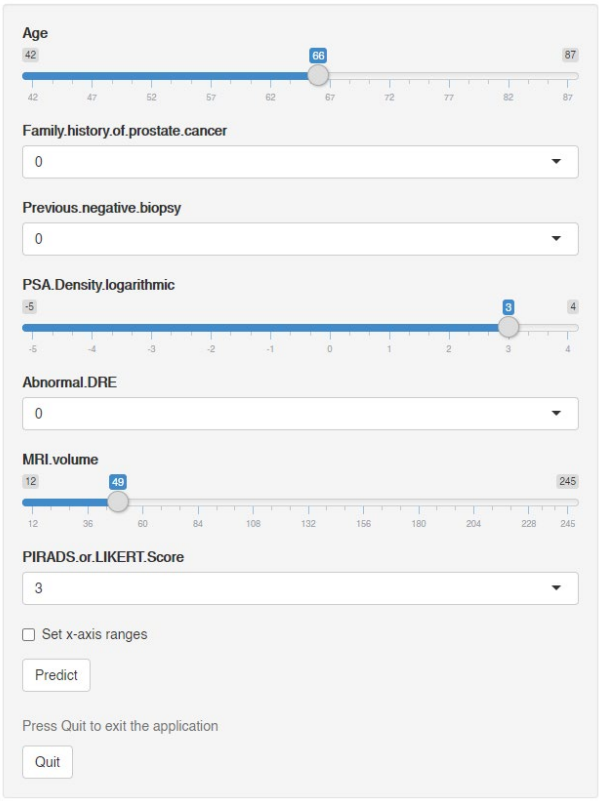


Figure 73: Screenshot of web version of the nomogram using the Shiny app.

11.4.6 External validation

The external validation of Model 1 for the primary definition (Gleason ≥ 3+4) and one secondary definition (Gleason ≥ 4+3) was performed in six international cohorts. The external validation was limited to these definitions due to lack of availability of maximum cancer core length in the external dataset which makes calculation of UCL/Ahmed 1 definition of clinically significant prostate cancer less reliable. A summary of the outcomes are shown in Table 61.

Table 61: Discrimination of the model in development cohort compared with external cohort

	Gleason ≥ 3+4	Gleason ≥ 4+3
Development	0.8230	0.8144
Cohort 1	0.8080	0.8170
Cohort 2	0.7951	0.7533
Cohort 3	0.8554	0.7844
Cohort 4	0.8257	0.8026
Cohort 5	0.8158	0.8066
Cohort 6	0.8229	0.8117

Data are AUC (Area under the curve) for each cohort

11.4.6.1 Gleason ≥ 3+4

The calibration plots for the external six cohorts showed that the model overestimated risk of Gleason ≥ 3+4 in the higher probability deciles (30%-80%). The model under estimated risk in 10-20% deciles. The Hosmer and Lemeshow goodness of fit was border-line for statistical significance (X-squared = 15.206, p = 0.05525) so recalibration of the coefficients and intercepts were performed and the improved recalibrated model is shown in Figure 77.

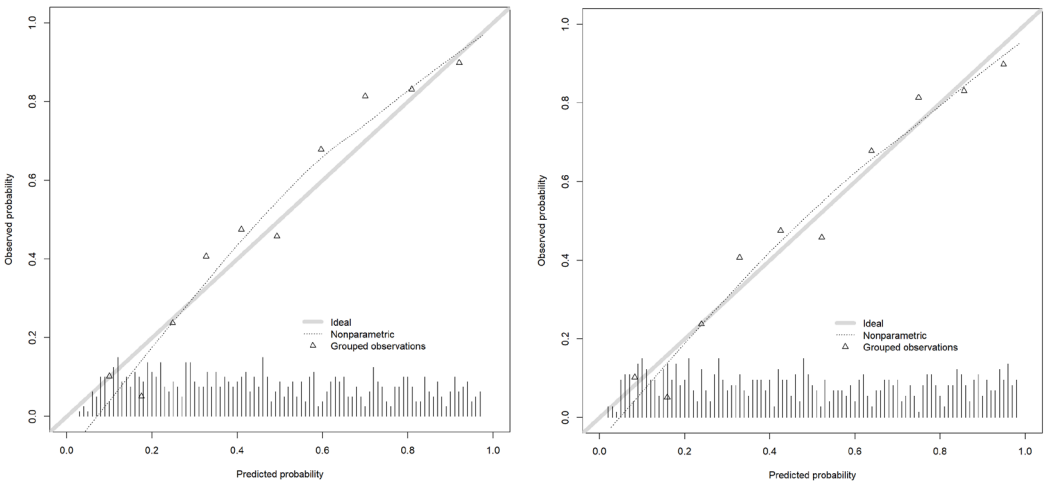


Figure 74: Calibration plots for the RAPID risk score in the validation cohort 6. (A) Calibration plot on model from development cohort (B) Calibration following a recalibration process showing improved calibration.

Application of the final model to the independent validation cohorts showed discrimination was 0.8080 in Cohort 1, 0.7951 in Cohort 2, 0.8554 in Cohort 3, 0.8257 in Cohort 4, 0.8158 in Cohort 5 and 0.8229 in Cohort 6. The AUC was ≥ 0.8 in all cohorts which is considered 'excellent'. Apart from Cohort 2 with AUC = 0.7951 which is 'acceptable'.

The decision curve analysis comparing the clinical usefulness of the model across each external cohort is shown in Figure 78. The model showed a net benefit in Cohorts 1, 3, 4 and 5 across multiple ranges of threshold probability. In Cohort 2, the model crossed the biopsy-in-all line suggesting that in this cohort at the range of threshold probability between 10-20% the model exhibited inferior performance to a biopsy-in-all strategy.

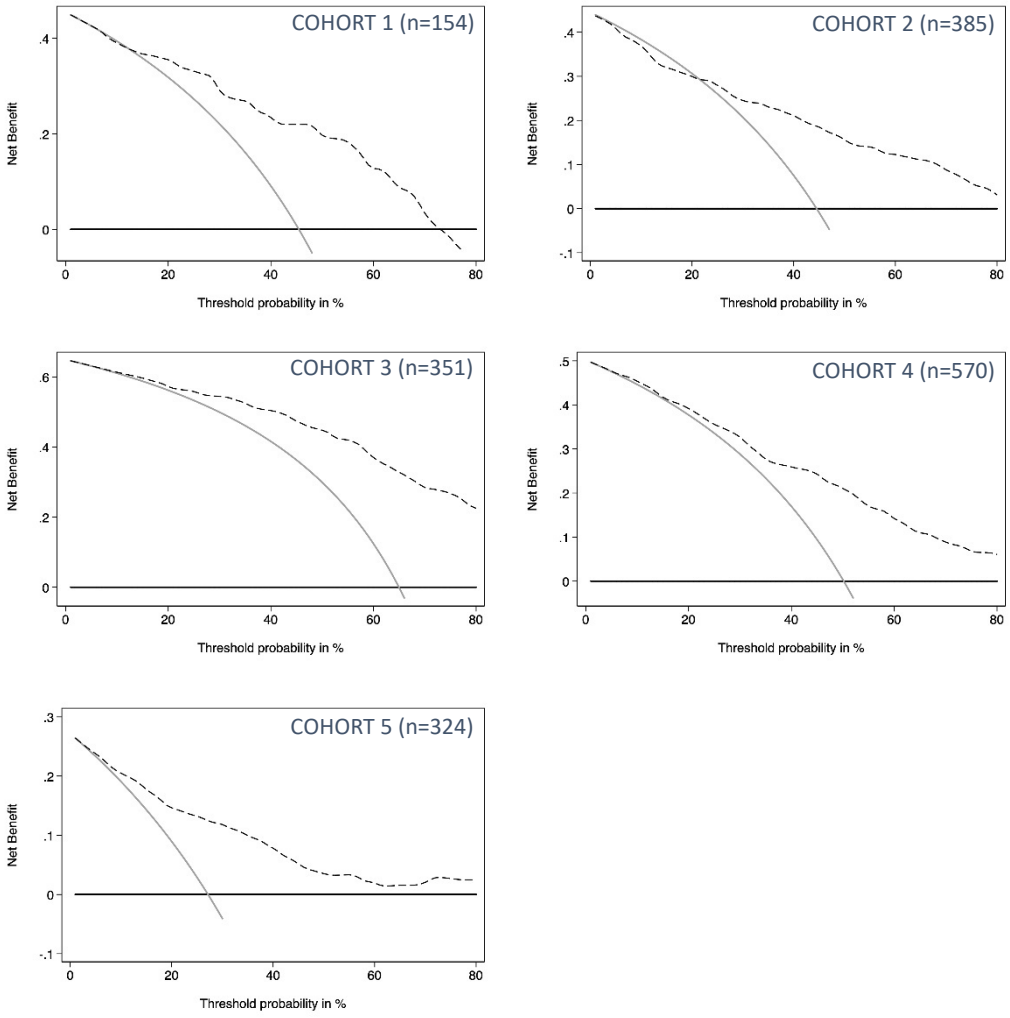


Figure 75: Decision curve analysis for Model 1 (Simple) across external validation Cohorts 1 to 5.

11.4.6.2 *Gleason ≥ 4+3*

The calibration plot showed good calibration at lower deciles and an underestimation of prostate cancer risk in patients >30% risk. A similar process of recalibration as for Gleason ≥ 3+4 was performed. Calibration improved as shown by the Hosmer and Lemeshow test which changed from 6.11 (p = 0.6342) to 8.79 (p-value = 0.3595) post recalibration.

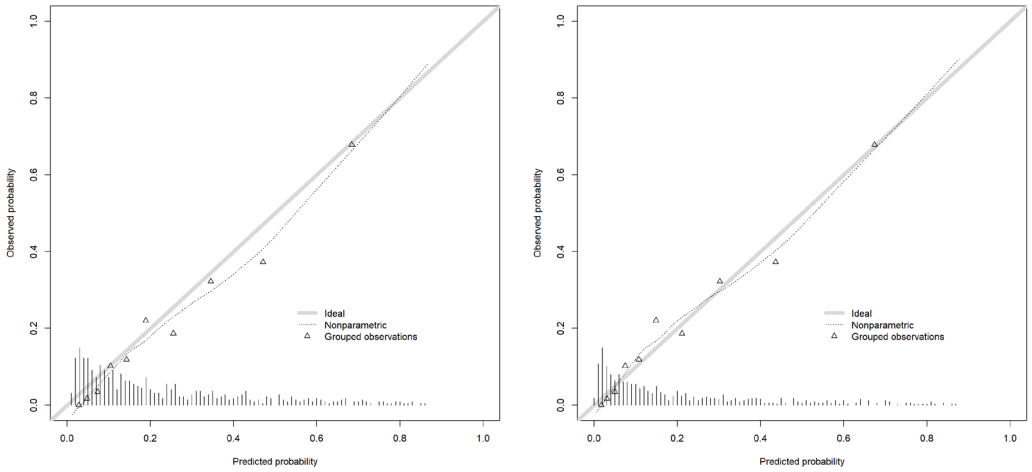


Figure 76: Calibration plots for the RAPID risk score in the validation cohort 6. (A) Calibration plot on model from development cohort (B) Calibration following a recalibration process showing improved calibration.

The discrimination of the model is slightly lower for Gleason ≥ 4+3 compared to the primary outcome, but this decrease was not statistically significant on DeLong's Test (p>0.05 for all cohorts). The AUC remained above 0.75 which is considered acceptable and remained similar across each cohort. The AUC in Cohort 1 was 0. 8170, in Cohort 2 was 0. 7533, in Cohort 3 was 0. 7844, in Cohort 4 was 0. 8026, in Cohort 5 was 0. 8066 and in Cohort 6 was 0.8117. The difference in AUC between primary and secondary endpoints might result from increased numbers of men with Gleason ≥ 3+4 in the external cohorts.

In decision curve analysis, the model showed a net benefit compared to the strategy of biopsy in all apart from Cohort 2 which showed a reduction in benefit in threshold probability range 10-15%. The remaining cohorts showed a positive net benefit for predicted probability thresholds after 5% compared to a biopsy in all or biopsy in none. For cohort 3 and 5, at lower thresholds (below 10-20%) there was no difference between using the risk model and conducting a biopsy in all patients.

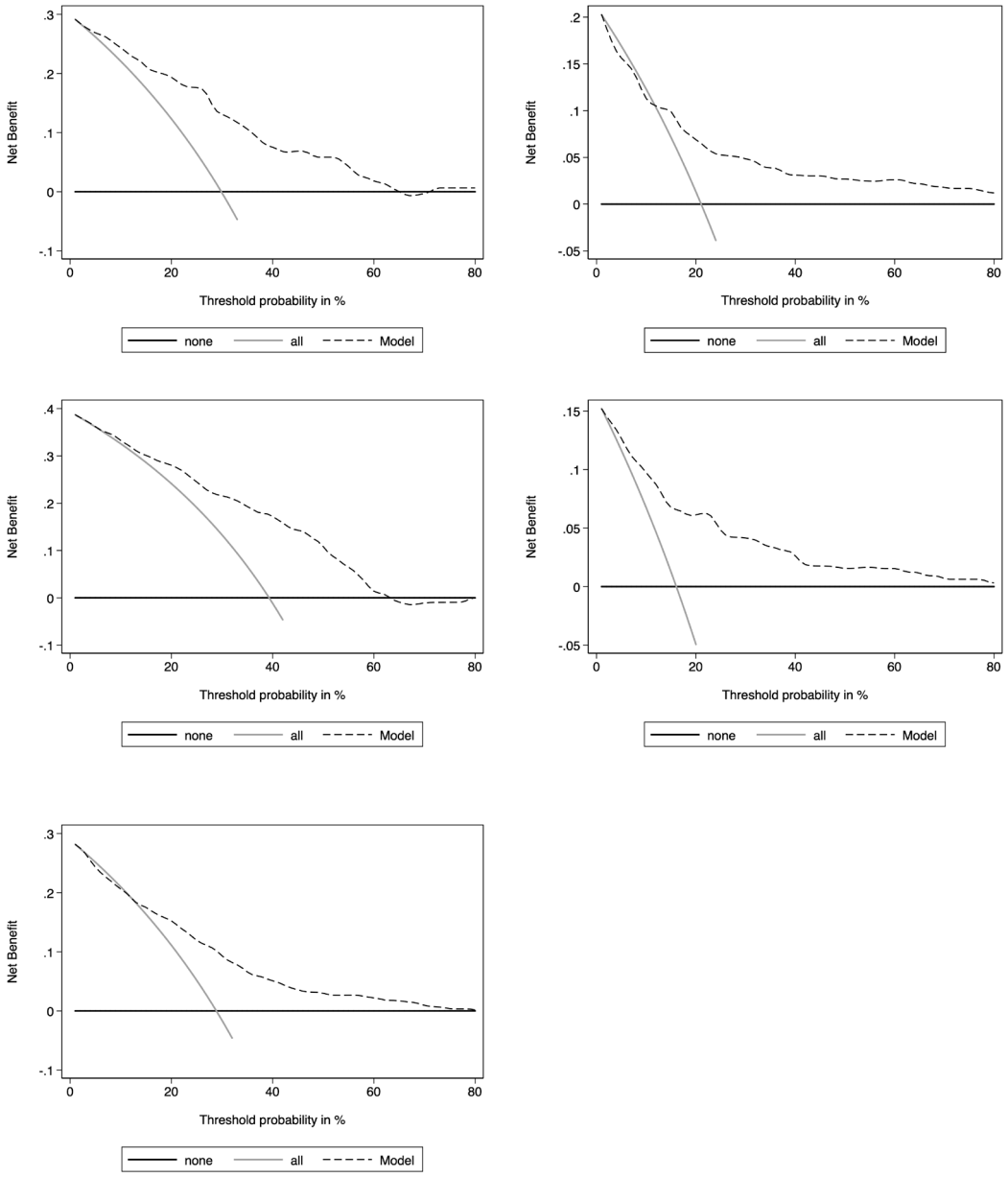


Figure 77: Decision curve analysis using the Rapid Risk Model for Gleason ≥ 4+3 in external cohorts 1-5. The decision curves estimate net benefit at a range of possible threshold probabilities. On the graphs the line that is highest over the widest range of threshold probabilities indicates the highest net benefit.

11.4.7 Comparison to existing models

The final risk model (Gleason ≥ 3+4) was compared to a alternative MRI risk score developed by Mehralivand et al [343] using Cohort 6 (external data). As shown in Figure 81, the AUC of the Mehralivand model had good discrimination (AUC 0.812, (95% CI 0.788-0.836) in the external data. In comparing the RAPID risk model to the Mehralivand model, our model had significantly higher discrimination based on DeLong's test[361] for correlated ROC curves (Z=2.2585 p ≤ 0.02392) with an AUC 0.823 (95% CI 0.8.00-0.846). The Mehralivand model had a similarly worse performance at the secondary definition threshold of Gleason ≥ 4+3.

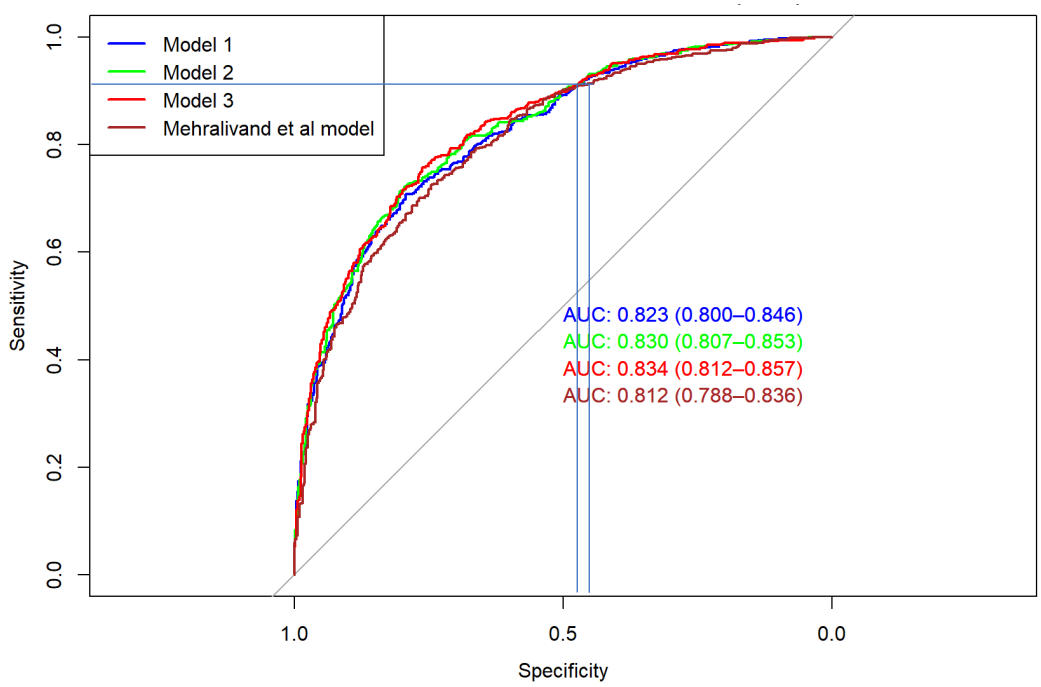


Figure 78: Performance of RAPID risk models and Mehralivand model in external validation cohorts.

Clinical utility analysis showed that at a sensitivity threshold 90%, the corresponding specificity for the most parsimonious model (Model 1) would be 47.9% compared to the Mehralivand model at 46.7%. This improvement in specificity is small but in a screening population will have an impact on the number of false positives. For example in a UK screening population where there are 7,859,000 men aged 50 to 70 years eligible for screening and an assumed prevalence of significant disease 2% as per the PCPT trial, this improvement in specificity corresponds to 92,424 reduction in false positive MRI scans requiring a prostate biopsy.

11.5 Discussion

11.5.1 Principle findings

This chapter aimed to develop a simple model that was able to differentiate benign and malignant MRI lesions across different MRI scoring systems and definitions for clinically significant prostate cancer. The basic RAPID risk model (Model 1) is a simple five-item score which provides a standardised tool for the prediction of clinically significant prostate cancer in men with a visible MRI lesion. In the context of an MRI screening pathway, this tool could support patients and clinicians making decisions regarding the need for prostate biopsy.

Model 1 was selected as clinical utility analysis showed that difference in net benefit per patient was relatively small compared to the less parsimonious models. This model has the advantage of being simple without compromising diagnostic accuracy or requiring numerous MRI variables. The co-variates included are ones which are easily available for men undergoing MRI including age, prior biopsy history, PSA, prostate volume and MRI score. To further facilitate easy access, I created an app of this model using the R language (available at https://rapidriskscore.shinyapps.io/RapidRiskScore1/).

A key strength of this analysis was externally validating the model on multiple datasets and comparing it to an existing scoring system. The model showed consistent performance in six external validation datasets from multiple countries using different MRI scoring systems and diverse reference standards. The primary model (RAPID RISK) demonstrated acceptable calibration between predicted and observed risk in external cohort 6. Although there was some miscalibration in the lower risk strata, meaning that as risk decreased below 20%, the model underestimated risk. However, after recalibration the estimates were superior. With further rounds of testing and recalibration in different cohorts, the model has potential to be useful in a wide range of clinical settings.

The RAPID risk model was developed in response to recent calls that we move towards a risk-stratified MRI pathway for prostate cancer diagnostics[364]. The model was designed to provide men with a visible MRI lesion with a personalised risk of significant prostate cancer across multiple definitions. In this chapter we examined the trade-off if the model was deployed as a decision tool in men with an MRI lesion who are considering avoiding a biopsy. The decision curve analysis suggested that the primary model is likely to be useful for men with a predicted risk of prostate cancer beyond 10%. Whether this probability threshold is useful clinically requires patients to balance their views on the risks of undergoing an unnecessary biopsy

with the prospect of missing clinically significant prostate cancer. If patients consider the risk of false negative (missing significant cancer) very heavily, then the model might need to yield a net benefit across predicted risk cutoffs below 10%. If the majority of patients have a similar risk tolerance then the model would not be useful in clinical practice. In this chapter, I do not speculate whether this model will be useful in informing the choice of patients about whether to undergo a biopsy. Instead this will be evaluated in a prospective clinical trial where patients are given results of the model to decide whether to have a biopsy or undergo observation.

11.5.2 Comparison with previous studies

Although there have been several models which have been developed that predict the risk of significant prostate cancer based on MRI score, not all have been externally validated and none have been designed specially for men with an MRI lesion. In comparing the RAPID risk model to a previously validated model developed by Mehralivand et al[343], I have shown that the RAPID risk model had a small but significant improvement in performance in the external validation cohort. The AUC was significantly higher in Delong's test showing that the RAPID risk model is more effective at discriminating benign and malignant lesions. If applied to a UK screening population utilising the model would lead to a reduction in 92,424 false positive MRI results.

Additional in the Model 2 (non-invasive) I include Afro-Carribean ethnicity which is a strong predictor of significant prostate cancer but absent from numerous alternative models due to lack of racial diversity in the development cohort. For example, the most common risk models which have been used come from the Prostate Cancer prevention trial (PCPT) risk calculator and the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk tool[365, 366].

However, the ERSPC included predominately white European men who underwent heavy screening and the PCPT included mainly healthy white men from North America who required a normal DRE and PSA ≤ 3 to be eligible for the trial. Neither included a racially diverse population in model development in contrast to the RAPID risk score where 12% of the cohort were of Afro-Caribbean ethnicity.

In addition, both these trials were completed over two decades ago, using historic diagnostic practices and specific eligibility criteria for the cohorts. In PCPT, men had annual PSA and DRE screening and the outcome was based on an end of study TRUS biopsy after seven years. Both cohorts utilised 10-12 core non-targeted TRUS biopsy which was graded using the

previous versions of the Gleason grading systems which reclassified certain groups into higher Gleason Scores. In contemporary practice screening for prostate cancer is rare and men are only likely to undergo an assessment for prostate cancer following an elevated PSA[40].

11.5.3 Limitations

This model has several limitations;

First, the primary model was chosen for its simplicity but this may lead to reduced discrimination and utility in clinical practice. While some modelling studies use a single method to define predictor variables, we used a complementary approach which balances objective statistics and clinical decisions when determining the variables for the model[367]. Including too many variables results in more bias, equivalent to inflated Type I error and there is an ongoing tension between maximising performance and useability of the model in real life practice. A general rule is that a simplistic and parsimonious model that has a good fit with the data is preferred[367]. For these reasons, the simple model was chosen as the preferred option, given it showed similar net benefit compared to the more complex models.

Second, the model was derived from observation data rather than a clinical trial. There is an inherent risk of inaccuracies in data collection, miscoding, and missing data in any large, multicentre clinical database. However, the RAPID Online data has been rigorously designed, standardised and validated so it is expected that the errors should be minimal. The pathway has standardised eligibility criteria and a priori defined datapoints which strength the design of the model. There is a benefit from using real-life data and RAPID Online represents a uniquely large and standardised observational data set including all patients undergoing the RAPID pathway at three large university hospitals in the United Kingdom.

Third, the pre-specified eligibility for RAPID may limit the use of the model in all men with an MRI lesion. In addition, in the RAPID cohort not all patients with an MRI lesion received a targeted biopsy. This may lead to verification bias at lower MRI scores when patients are more likely to not be biopsied.

Fourth, the model has been developed and validated in a diagnostic rather than screening setting. The model needs to be validated within a screening population and the IP1-PROSTAGRAM cohort was considered. However, due to the low event rate for the primary

outcome in IP1-PROSTAGRAM, we could not perform a robust external validation on this cohort.

Fifth, the simple model did not include a number of variables which have been shown to be predictive for prostate cancer, particularly digital rectal examination. This was excluded from Model 2 and 3 on the basis of missing data and to reflect the real-life scenario that not all men who progress in an MRI pathway have a DRE. The model was designed to include only predictors routinely available to men after completing an MRI.

Sixth, the model was restricted in use of predictor MRI variables and type of MRI score. A composite MRI score was used (highest PI-RADS or Likert) due to high levels of multi-collinearity between MRI variables. I attempted to overcome this limitation by validating the model on cohorts which used multiple different scoring systems.

Last, although I developed a nomogram, the model was validated in the form of a logistic regression equation which limits its ease of use. The next logical step following this analysis is to prospectively test the impact of this model in clinical practice in addition to further external validation in different cohorts.

11.6 Conclusion

This chapter has developed and validated a contemporary risk prediction model to differentiate benign and malignant MRI lesions. The new scoring system showed good discrimination in six independent validation cohorts of men with a visible MRI lesion. It is yet to be validated in a screening population but if it shows similar performance in this setting it would allow a more personalised screening pathway. The decision support tool which has been published online has been designed to reduce patient anxiety and uncertainty, improve decision quality and increase acceptance of initial surveillance of low risk MRI lesions. The impact of this decision support tool will be evaluated in a prospective clinical trial based on the risk model developed in this chapter.

Chapter 12 – Discussion

12.1 Overview

This final chapter provides an overview of my thesis and an evaluation of its implications for future research. A summary of the key findings of each chapter is provided followed by important limitations encountered. Finally, the chapter finishes with some concluding remarks and recommendations.

12.2 Summary of findings

The purpose of this thesis was to understand whether MRI has a role as a screening test and to explore methods of integrating it into the pathway for screening of prostate cancer. The thesis was grouped into three parts which will be summarised in turn:

12.2.1 Part 1: Literature background

The merits of prostate cancer screening have been subject to widespread debate. Much of the recent debate has been inextricably linked to the use of PSA which has received a high level of scrutiny but has been contentious as a screening test due to a combination of under- and overdiagnosis of the disease. Chapter 2 presented some of these key PSA screening studies, highlighting that although PSA screening may reduce prostate cancer specific mortality, at a threshold of ≥3ng/ml, it is not clear that the mortality benefits outweigh the risks from false positives and overdiagnosis of insignificant prostate cancer.

The potential for MRI as an alternative method of screening was considered in Chapter 3. The selection of MRI was motivated by its success as a diagnostic test in secondary care where it has been shown to be a highly sensitive test for the detection of clinically significant cancer while minimising unnecessary prostate biopsy. As noted in the initial chapters, at the start of this thesis there was virtually no evidence relating to how MRI might perform as a screening test. Due to this paucity of existing evidence, a narrative review was completed to synthesise a broader range of themes, including exploring the challenges related to MRI as a screening test drawing on lessons from other image-based screening programmes.

The findings of this chapter suggested that while MRI could have certain characteristics attractive for population-based screening, there would be challenges related to its low specificity, high cost and limited availability. Certain issues might be mitigated by recent

technological advancements in MRI hardware and developments of shorter functional sequences. Therefore, this chapter concluded by recommending that the first step towards a screening MRI would be an evaluation of a shorter MRI protocol.

12.2.2 Part 2: Evaluating the performance of MRI

The second part of this thesis commenced with evaluating the performance of two alternative shorter MRI protocols with the aim of identifying the MRI sequences which could form the technical basis for a fast MRI for screening. In Chapter 4, we identified that the combination of T2W and DWI sequences provided a balance of maximum diagnostic accuracy while reducing excessive biopsies.

Having identified the MRI sequences for screening, the next step was to evaluate this fast MRI protocol in the general population at both score thresholds (MRI Score ≥ 3 and MRI score ≥ 4). Chapter 5 explained the recruitment process for the IP1-PROSTAGRAM trial which was the first clinical trial to compare the performance of a fast MRI and PSA as exclusive screening tests for prostate cancer. The study used a paired screen-positive design to compare the new fast MRI and PSA, in which a biopsy was recommended in the presence of any screen-positive test and participants remained blinded to the indication for biopsy until the biopsy procedure was completed. This design has been used in other screening studies for prostate cancer[249] and addresses the common methodological challenge for screening studies where the reference test cannot be carried out on all participants due to ethical, practical and cost reasons.

The primary results were presented in Chapter 6 which showed that an MRI score ≥4 found more significant cancers than PSA without increasing the harms from additional prostate biopsies and overdiagnosis of insignificant disease. The effects of verification bias were considered in Chapter 7 where a variety of methods were used to correct the bias and provide point estimates for sensitivity and specificity for each test.

Chapter 8 evaluated a secondary outcome of acceptability of fast MRI compared to PSA. Acceptability is fundamental to the successful delivery of a high-quality screening programme although it is often overlooked or not robustly evaluated within clinical trials. It is particularly important for screening programmes as they rely on high levels of uptake and acceptability of screening tests. Both PSA and fast MRI had high levels of acceptance among participants. Fast MRI was the preferred overall test of participants although PSA had a slightly lower level of overall burden. The analysis of predictors of MRI burden indicated that

the most important considerations in overall burden of a screening test was related to the initial perception of the test rather than background or procedural factors.

12.2.3 Part 3: Improving the performance of fast MRI

The final part of this thesis was committed to methods to improve the performance of fast MRI and minimise some of the challenges associated with the test. The first issue addressed was an alternative approach of combining PSA and fast MRI in a screening pathway. Different pathways of PSA and fast MRI were compared in Chapter 9 using a variety of definitions for significant disease to identify the optimal combination for a new screening pathway. This identified a pathway which combines PSA ≥ 1ng/ml and MRI score ≥ 4 as a better combination of tests for triaging men for a biopsy than a single independent screening test. This approach has the advantage of preserving the simple and reproducible PSA test while not impacting diagnostic accuracy.

Chapter 10 provided empirical evidence to support the deliverability of such a PSA and fast MRI screening pathway. This chapter utilised the RAPID registry which was funded, developed and implemented as part of this doctoral programme of work. The RAPID pathway is similar to a structure which could be set up to deliver MRI within an organised population screening programme. A historical cohort allowed estimation of a counterfactual following implementation of the pathway, against which the new pathway outcomes could be compared. The results suggest that the RAPID pathway could allow high-volume, rapid diagnosis of men who require screening for prostate cancer.

Chapter 11 utilised the RAPID registry to develop a risk model in order to improve upfront risk stratification and reduce the number of prostate biopsies in men with low risk lesions. Multivariable logistic regression models using the backward stepwise method were used to develop a risk prediction model. The final model performed well in differentiating benign and malignant MRI lesions across a range of definitions for clinically significant cancer. It also maintained its predictive ability when validated in multiple, international, external cohorts with diverse reference standards and MRI scores. Moreover, the RAPID risk model's discriminatory performance was superior to two other published MRI risk models in the external validation cohort.

12.3 Limitations and Solutions

In each chapter of this thesis, limitations of the relevant study have been discussed. In this section, additional limitations that cross-cut the thesis will be highlighted along with potential solutions:

12.3.1 Impact on disease-specific mortality

It is acknowledged that this thesis did not attempt to evaluate the impact of fast MRI on prostate cancer-specific mortality, which is the ultimate test of the efficacy of a screening test. Instead the findings are only relevant to the diagnostic performance of fast MRI in a screening setting. There are mechanisms by which one could hypothesise that the improved detection of significant prostate cancer by MRI could translate into a reduction in mortality. However, it is important to emphasise that such a conclusion has not been proven in this thesis.

Indeed, if one attempted to evaluate mortality impact by longitudinal follow up of the IP1-PROSTAGRAM cohort this would be limited by lead time and length time bias. For example, it is possible that fast MRI may be detecting the same cancers as PSA but earlier in their natural history. PSA may have still been able to detect these cancers at a favourable stage in which case fast MRI's early detection would confer minimal mortality advantage (lead-time bias).

Equally it is possible that the additional cancers detected by fast MRI may be less aggressive and have a better outcome than PSA detected cancers (length time bias). In IP1-PROSTAGRAM there were no cases of Gleason ≥ 4+3 detected, likely due to the sample size and low risk population. It is this higher grade Gleason ≥ 4+3 which has been shown to be most predictive of prostate cancer specific mortality in longitudinal cohorts of watchful waiting with 29 years follow up[38] . Without any cases of Gleason ≥ 4+3, the finding that fast MRI may detect more cancer than PSA relies on conservative definitions such as Gleason ≥ 3+4, UCL/Ahmed 1 and UCL/Ahmed 2. It is possible that these definitions may be too conservative[177] and that fast MRI may be increasing overdiagnosis by detecting cancers which were not going to cause morbidity or mortality.

Ultimately, these issues can only be addressed with a randomised controlled trial evaluating prostate cancer specific mortality. Such a trial would remove the effects of lead and length-time bias although it would require a prolonged follow-up period due to the long natural

history of screen-detected prostate cancer. Longitudinal follow-up of the IP1-PROSTAGRAM cohort will not provide a definitive answer to disease-specific mortality due to the non-randomised design. Instead, an RCT is needed with a similar design to CAP, ERSPC and PLCO trials using an MRI screening pathway.

12.3.2 Low prevalence of high-grade disease

The low event rate of significant cancer in IP1-PROSTAGRAM is an important limitation for chapters comparing the diagnostic outcomes of fast MRI and PSA. This low event rate was anticipated in the trial design, and the sample size for IP1-PROSTAGRAM was determined by the primary outcome which was selected to be realistic and practical to complete within this doctoral thesis timeline.

However, the sample size meant that the study was not powered to compare cancer detection rates, meaning that diagnostic accuracy calculations (Chapter 7) and pathways analysis (Chapter 9) are exploratory in nature and further work is needed to confirm the findings of these chapters. Such a comparison will require a larger diagnostic accuracy study to compare sensitivity and specificity of PSA and MRI as screening tests. A follow-up study could be designed, powered on a hypothesis that a fast MRI has a higher sensitivity than PSA for detection of significant prostate cancer as a screening test. The sample size for this analysis can be calculated using the estimates from IP1-PROSTAGRAM using the method described by Alonzo et al[368] estimating 80% power and a two-sided significant threshold of 5%.

Using this method, it is estimated that a sample size of about 2,320 men would be needed. This was determined using the uncorrected sensitivity of MRI (TPR_{MRI}) 64.7% and 41.2% for PSA (TPR_{PSA}) from Chapter 7. The proportion of men with significant cancer who tested positive on both tests (TPPR) was 35.2% based on the IP1-PROSTAGRAM outcomes (6 of 17 significant cancers were positive on both tests). If 2,320 men were recruited, it would be expected to find 51 cases of clinically significant prostate cancer given a prevalence of significant prostate cancer of 2.2% from the Prostate Cancer Prevention Trial (PCPT)[86, 158]. This would generate 33 cancers in the MRI group and 21 cancers in the PSA group which would be sufficient to compare sensitivity and specificity.

12.3.3 Impact of MRI interobserver variability

A key feature for a screening test is that the results must be reproducible meaning it must provide consistent results when performed across diverse centres and interpreted by different clinicians. However, in Chapter 6 I found that fast MRI had a high interobserver variability when measured by kappa as fair/low.

Computer-aided detection (CAD) or Artificial Intelligence (AI) systems may be a potential tool to reduce interobserver variability and improve radiological reporting capacity. CAD/AI systems actsas a supplement to human readers and marks potential areas of concern so the radiologist can decide if the area warrants further investigation. There has been widespread use of CAD/AI systems in breast cancer screening programs particularly in the United States' Medicare population where it is estimated it is used in 74% of screening mammograms. There is extensive evidence that the system can improve the sensitivity of mammography and improve radiology reporting workflow. CAD/AI systems are being investigated in other image-based screening modalities particularly CT-colonography with encouraging results and there are similar CAD/AI systems available for prostate MRI and in the early stage of evaluation which could be validated using the IP1-PROSTAGRAM cohort.

An additional solution to reduce interobserver variability is double readings, as occurs in the majority of European breast cancer screening programmes. A double reporting strategy should be considered in future screening trials of fast MRI, particularly given that PSA, as the comparator test, has the advantage of being more reproducible. In IP1-PROSTAGRAM resource constraints dictated that the double reporting strategy was limited to 20% of the original MRI scans. It is recommended that future trials have independent double interpretation with arbitration of every MRI scan.

It should also be acknowledged that for the radiologists reporting within IP1-PROSTAGRAM this was their first experience reporting a prostate MRI without secondary clinical information such as PSA and digital rectal examination findings. It is well-recognised that experienced readers outperform novices when interpreting complex scans and the novel nature of IP1-PROSTAGRAM meant that no radiologist had experience in this area prior to the study. There remains considerable scope to improve training and interpretation of a screening prostate MRI prior to any further trials.

Previous image-based screening tests such as mammography and low dose CT have improved interobserver variation by use of validated criteria during interpretation. Although the

radiologists were reporting using validated scoring systems, neither had been designed for a screening population. Both scoring systems, Likert and PI-RADS, have been developed for a secondary care (high-prevalence) setting so may not have been appropriate for a screening setting.

For example, in IP1-PROSTAGRAM the primary source of variability was MRI Score 3 (indeterminate lesions) which we do not propose to include within the definition of a screen-positive fast MRI. It is reassuring that other measures of interobserver variability showed higher level of agreements for MRI Score ≥ 4. It is also possible that the kappa values may have been influenced by a statistical paradox which is well reported in the literature[369, 370] and shows that kappa may not be the optimum measure of interobserver agreement in a low prevalence setting. The kappa metric has been the most common method of reporting interobserver variability in previous clinical trials of MRI but these were conducted in high prevalence settings such as radical prostatectomy cohorts[371] or men referred with elevated PSA levels. However, when the hypothetical probability of chance agreement among raters is high, kappa can be paradoxically low even with high levels of observed agreement (e.g. 70.5%).

In IP1-PROSTAGRAM, I attempted to control for this statistical paradox by stratifying the randomisation of scans selected for double reporting. However, this may not have been sufficient to correct for the prevalence bias particularly in the score 4-5 threshold. Instead, using measures which are not affected by skewed distributions such as the AC1 coefficient, leads to a higher agreement of 0.51 (moderate) for MRI score 4-5 which is more reassuring than the fair/low kappa results.

12.3.4 Cost effectiveness

The final limitation which cross-cuts this thesis is the fact that a cost-effectiveness analysis of fast MRI was considered beyond the scope of this thesis. Clearly, a formal cost-effectiveness analysis must form part of future screening trials and it is acknowledged that even a short, non-contrast 10-minute MRI protocol has considerable additional costs compared to PSA. The cost on the NHS tariff for a standard non-contrast prostate MRI would be £108 per scan including reporting[188] in comparison to a PSA test at £4[189].

A comparison of the cost of fast MRI to other cost-effective screening tests is shown in Table 62. Although not a formal cost-effectiveness analysis, it serves as a proxy for how MRI (£108) compares to other cost-effective screening tests such as LDCT (£69), CT colonoscopy (£97) or

flexible sigmoidoscopy (£304). One must consider that the process of evaluating and potentially incorporating fast MRI into the prostate cancer screening pathway will take many years so it is reasonable to assume that the cost, acquisition speed and image quality of MRI will further improve. Over the last decade the cost of MRI has already reduced by 40% and this trend is expected to continue[372]. A comparable example has occurred with CT for lung cancer screening which was not cost effective until the development of low-dose helical CT with a rapid image protocol which could perform scans in seconds at a similar cost to mammography[373].

Table 62: Comparison of screening test sensitivity, specificity and cost

Screening Test	Sensitivity	Specificity	Cost
Group 1: Reduced sensitivity and lower cost			
PSA >3ng/ml (prostate cancer)	58%[86]	82%[86]	£4[a]
Stool tests (FIT) (colorectal cancer)	32%[374]	85.8%[374]	£8.09[375]
Mammography (breast cancer)	40%[285]	95%[285]	£33.50[190]
Chest X-ray (lung cancer)	73%[376]	91%[376]	£17.40[b]
Group 2: Improved sensitivity and higher cost			
PROSTAGRAM (PI-RADS ≥ 4)[b]	78.0%	91.6%	£108[c]
Endoscopy/Imaging			
- Flexi Sig	83%	60%	£304[b]
- Colonoscopy	100%	43%	£451[b]
- CT colonoscopy	97%[374]	40%[374]	£97[b]
Breast MRI (breast cancer)	71%[285]	90%[285]	£249.60[190]
Low dose CT (lung cancer)	94%[376]	73%[376]	£69[b]

[a] NHS National Schedule of Reference Costs 2016/17: DAPS04 + DAPS08

[b] For UCL/Ahmed 1 according to multiple imputation method in Chapter 7

[c] NHS National Tariff 2019/20. Unit cost includes cost of reporting.

CT colonoscopy assumed equivalent to 3 areas with CT contrast as in Porté et al[377]

Another factor to consider is that the cost-effectiveness of screening includes many factors beyond the unit cost of the test. It will depend on the way in which MRI screening is implemented, specifically regarding the selection criteria, screening interval, diagnostic follow-up of equivocal lesions and treatment (Figure 82). An equivalent example is flexible sigmoidoscopy or colonoscopy screening which is delivered at 5 to 10 year intervals and has

been shown to be cost-effective despite having a significantly higher operational cost than the faecal occult blood test[196, 378, 379].

If a similar approach was taken for MRI screening using a 10 year interval, men could be offered MRI screening twice at between 55-60 years then 65-70 years. Annually, there are approximately 460,000 men entering the 55-60 year age group and 340,000 men entering the 65-70-year age group in the UK[380]. The estimated annual throughput for a single MRI scanner would be 8,096 men per year based on a 15-minute total procedure time, 8 hours per day and 253 working days per year. In total, 99 MRI scanners would be required to deliver such an MRI screening programme in the UK. This equates to approximately £79.2 million capital expenditure with estimated ongoing costs between £14-169 million per annum based on costs of existing AAA or breast cancer screening programmes[381].

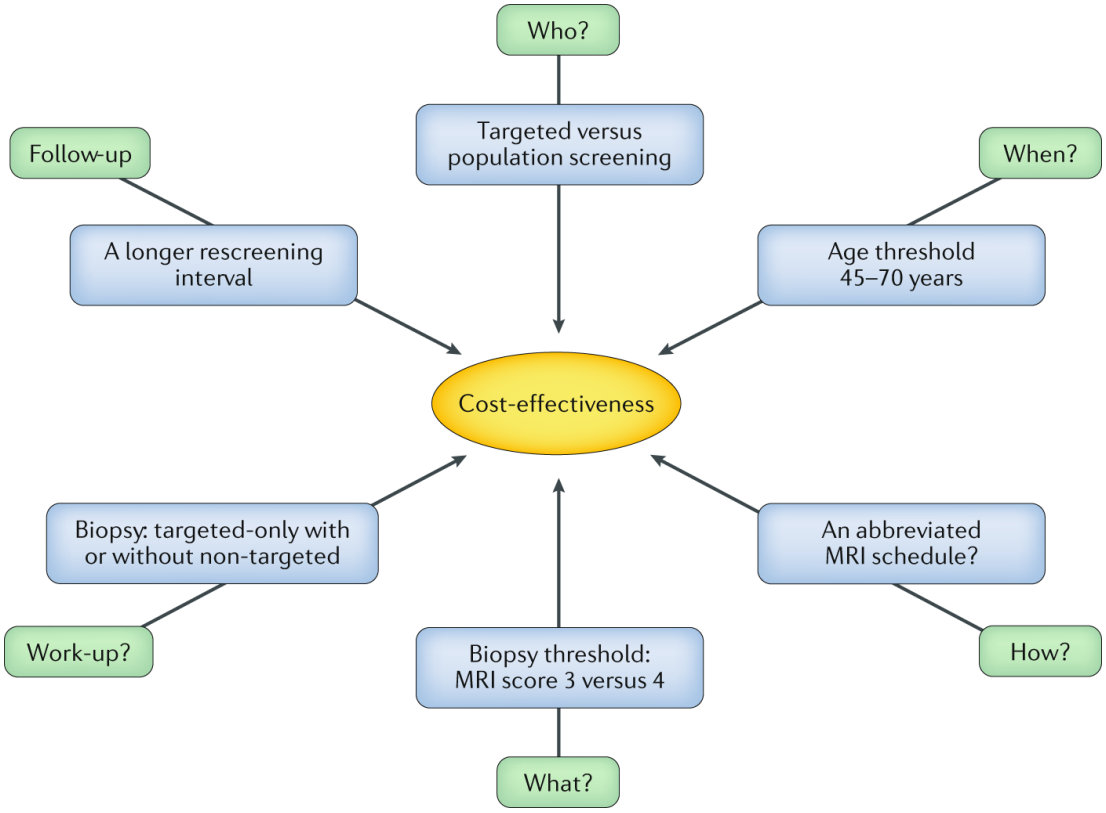


Figure 79: The cost-effectiveness of an MRI screening programme will be dependent on a range of factors including the following: whether MRI can be targeted to a certain population; what the most appropriate age range is for MRI screening and other factors.

Whether this expenditure represents a 'cost-effective' screening programme will depend on information which can only be acquired from a large randomised controlled trial. Ultimately the cost effectiveness of a screening programme should be determined on the cost per year of life/QUALY gained; for example, in PSA screening the estimated cost is $73,000 per QALY gained[382] but this requires empirical outcomes of the number of deaths prevented,

treatments and life-years gained from screening. At present any estimates of cost-effectiveness would be based on inaccurate assumptions of the potential benefits and harms from MRI screening. To acquire this information requires further investigation and future studies are planned to follow the findings in this thesis.

12.4 Contributions to the literature

The studies in this thesis have generated a number of new findings as well as providing support for existing research. A summary of key contributions to the literature are:

12.4.1 Development of a fast MRI for screening

This thesis has described the development and evaluation of a fast MRI protocol which could be appropriate for screening. Chapter 4 developed this shorter protocol using data from the PICTURE trial and showed that a biparametric MRI can decrease image acquisition times to around 10 to 15 minutes while yielding a similar diagnostic performance to full MRI.

The findings from this chapter provide further support to the accumulating evidence in favour of biparametric MRI consisting of only T2w and DWI sequences[125, 131, 137, 198, 278, 383]. Since publication the outcomes from this chapter have been included in meta-analysis by other research groups comparing biparametric and multiparametric MRI[384]. The outcomes of these meta-analysis support the findings in this thesis showing no significant improvements in sensitivity by omitting DCE.

12.4.2 Defining a screen-positive MRI

A new screening test requires setting an optimal threshold and Chapter 6 has established an optimal threshold to denotate a screen-positive MRI. In this chapter, fast MRI was evaluated as an independent screening test and the results showed that an MRI score ≥ 4 was the most appropriate threshold. These findings will provide justification when selecting the threshold to define a screen-positive MRI in future studies of MRI screening.

IP1-PROSTAGRAM was the first clinical trial to evaluate the performance of a fast MRI on a large scale in a screening setting. At the time of writing, IP1-PROSTAGRAM remains the largest population-based screening study to compare the outcomes of both PSA and MRI as screening tests conducted independently and blindly reported. Based on the outcomes of IP1-PROSTAGRAM, there are plans for future clinical trial by the Imperial Prostate research

group which incorporate the results of this thesis using the cut-off MRI Score ≥ 4 as the threshold to denote a screen-positive test.

12.4.3 Assessing the acceptability of MRI

Chapter 8 provided the first evidence on the extent to which fast MRI was acceptable as a screening test. In this chapter it was shown that the overall burden from fast MRI was low and pre-test expectations were the key determinants of burden. These results also highlighted that there is further work needed to improve the perceived burden of fast MRI given that it was marginally higher than PSA particularly in the domains related to anxiety, burden and embarrassment.

There are possible improvements which could be made to the IP1-PROSTAGRAM MRI protocol to further improvement acceptability. Possible areas of improvements include:

1. Reduction in image acquisition time: Shorter protocols which could reduce the amount of time patients spend inside the scanner. There have been recent studies which have evaluated new protocols which have further reduced scanning time to five minutes without any reported impact on diagnostic performance by including only axial T2w and accelerated DWI sequences[144]. These shorter protocols warrant further investigations for screening.

2. Oral administration of hyoscine butylbromide (Buscopan): The IP1-PROSTAGRAM MRI protocol included an intramuscular hyoscine butylbromide injection to reduce bowel peristalsis. This may have increased anxiety and burden levels prior to the scan and an alternative approach could be the use of oral hyoscine butylbromide given 30 minutes before the scan or no anti-motility drugs at all.

3. Wider bore and open scanners: When patients are anxious due to claustrophobia they generally prefer open and larger bore scanners[385]. Newer, wider bore and shorter length MRI scans will reduce claustrophobic reactions to MRI and are increasingly more prevalent. There are also many ongoing developments in MRI such as the introduction of pulse sequences to reduce noise which is an additional concern for patients undergoing MRI[386].

12.4.4 A contemporary re-evaluation of PSA as a screening test

The design of IP1-PROSTAGRAM was unique as biopsies were performed across all levels of PSA even in range 0-1ng/ml and the diagnostic accuracy calculations corrected for verification bias were reported in Chapter 7. This was one of the first modern studies to re-evaluate the diagnostic accuracy of PSA since the PCPT trial which was conducted in the era of less accurate and non-targeted TRUS biopsy.

Recent contemporary screening studies evaluating new biomarkers have applied a cut off to determine biopsy such as PSA ≥ 1ng/ml. This approach biases against PSA by removing a cohort where PSA has the highest level of diagnostic accuracy. An example of this was the STHLM-3 study where the biomarker was only used in men with PSA ≥ 1ng/ml [387]. Meanwhile technological improvements in biopsy and imaging has meant that IP1-PROSTAGRAM benefited from having access to the latest image fusion ultrasound targeted biopsy facilities as the reference standard.

The findings of this thesis, calling into question using PSA ≥ 3ng/ml as a threshold for screening, are in line with the results of the Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) which found that 36% of men who die from prostate cancer had a one-off PSA of less than 3ng/ml[67]. A baseline PSA of less 1ng/ml in men at age 60 has also been shown to be to be associated with an extremely low risk of prostate cancer specific mortality at 15 years in longitudinal cohort studies[388] and in PCPT[54] where only 7.2 of high risk cancers were identified in men with PSA > 1ng/ml. These studies provide further support for a combined pathway using PSA ≥ 1ng/ml as a method of stratifying men into groups of higher long-term risk of prostate cancer mortality.

12.4.5 A multi-modal screening pathway

Rather than replace PSA, Chapter 9 showed that the optimum approach might involve combining fast MRI and PSA in a multi-modal screening pathway. Such an approach offers the potential to maximize the positive performance characteristics of both tests while improving the feasibility and cost-effectiveness of the proposed screening programme. At a threshold of ≥1.0 ng/ml, the sensitivity of PSA for significant disease was >90%, and the risk of missing significant disease was low when combined with MRI Score ≥ 4.

These results provide a practical method of incorporating fast MRI into a screening pathway rather than a single independent test which would be challenging to deliver due to the high

cost and limited capacity of MRI. A pathway which combines PSA ≥ 1ng/ml and MRI Score ≥ 4 appears to provide the optimal balance between false positives, overdiagnosis and detection of significant disease. With respect to the number needed to screen (NNS), this chapter estimated that using Pathway 8, 41 men would be needed to be screened to diagnose one case of clinically significant prostate cancer. This compares favourably to the NNS in intervention arms from the first round of PLCO[389] at 143 and ERSPC Rotterdam[390] at 55 and CAP at 47[67]

The feasibility of delivering such a pathway is an important question from a healthcare policy perspective and was explored in Chapter 10. This showed that a one-stop pathway for screening could be acceptable to patients and it provided useful information to inform ongoing national healthcare decisions on reforming the prostate cancer diagnostic pathways for faster diagnosis.

A key consideration in the use of this pathway is the management of equivocal (PI-RADS/ Likert 3) lesions. Indeterminate lesions were the most common screen-positive result in the IP1-PROSTAGRAM study and had the highest level of interobserver variability. In breast cancer, efforts to reduce the false positive rate have led to the widespread acceptance of short-term follow-up algorithms of probably benign (BI-RADS 3) lesions for diagnostic mammogram. It is possible that a similar follow-up strategy could be replicated for MRI screening where men with equivocal MRI lesions are re-screened at an earlier interval.

Whether such a follow-up strategy is appropriate will likely be a source of debate in the literature given that IP1-PROSTAGRAM still found a reasonable number of significant cancers in men with indeterminate lesions. For breast cancer screening a mammogram is scored as probably benign (BI-RADS 3) when the chance of malignancy is less than 2%[391]. However, the PPV for equivocal lesions was 10.3% and 11.8% for PI-RADS 3 and Likert 3 respectively, which raises the question of whether short term follow-up is appropriate in this group.

An example of how this pathway could be designed is shown in Figure 83. In this pathway equivocal lesions are followed up at two years while screen-negatives are re-screened after 5-10 years. Stage 1 in this multi-modal pathway recommends a blood based biomarker which could be PSA ≥ 1ng/ml or a new marker, such as STHLM3, which may be proven more effective than PSA in the future. The only requirement of the biomarker is that the threshold is set to maximise sensitivity.

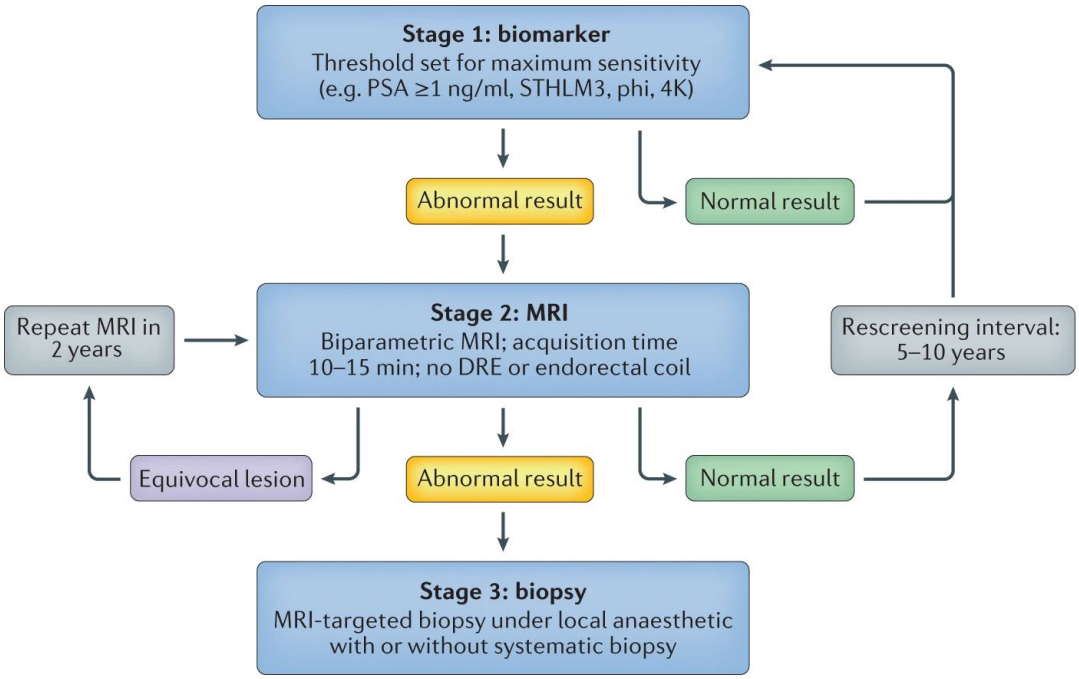


Figure 80: An example of a multi-modal screening pathway that combines a blood-based biomarker with a fast MRI.

12.4.6 Development of the RAPID risk model

To address the challenge of equivocal MRI lesions we looked at additional strategies to identify men at highest risk of clinically significant cancer who would warrant immediate prostate biopsy. The RAPID risk model was developed in Chapter 11 in order to further risk stratify these indeterminate lesions and to improve targeting of screening to those at highest risk of significant prostate cancer. The model has been validated across multiple cohorts in different clinical settings, a range of MRI scanners and different scoring systems which increases it generalisability[350-352]

It has been made available online at https://rapidriskscore.shinyapps.io/RapidRiskScore2/. If the model is validated in an MRI screening cohort it could be a useful tool to allow improved explanation of the risk of significant prostate cancer for men who are found to have an MRI lesion during prostate cancer screening. The first step is to evaluate the model in a prospective clinical study to determine biopsy decision. This clinical trial is due to commence recruitment in the near future.

12.5 Future studies and recommendations

The findings of this thesis have generated many questions: How should a screening pathway which includes MRI be designed? Could less frequent MRI screening intervals allow delivery of cost-effective MRI while maintaining high diagnostic accuracy? What definition of significant prostate cancer should be the primary target for MRI screening? Are the existing MRI scoring systems appropriate for a screening population?

Before fast MRI can be considered for screening, such questions need to be explored following a similar process to that required when extending the role of mpMRI to a pre-biopsy setting, which necessitated the development of new reporting systems, refining inclusion criteria and robust clinical trials.

A summary of the key areas for further exploration include:

12.5.1 A multi-centre randomised controlled trial of multimodal screening

Chapter 9 provided initial evidence that a combination of PSA and a fast MRI might provide a practical and accurate method of delivering MRI screening. In this pathway men are risk stratified to receive a fast MRI based on the initial screening results. Such a pathway would need to be evaluated in a large randomised clinical utility study powered to detect differences in cancer detection rates. The intervention group would follow this risk-adapted screening pathway while the control group would follow current standard NHS practice as issued by Public Health England. Based on the findings from this thesis a similar study is planned within the Imperial Prostate research group and is expected to invite 100,000-120,000 men in the community to participate via their GP practices.

12.5.2 Assessment of tumour behaviour within the IP1-PROSTAGRAM cohort

The challenge for clinical trials evaluating new screening tests is to account for contemporary and future shifts in prostate cancer prognostication. There is limited consensus on the definition of clinically significant prostate cancer which should be targeted by screening. Currently, the most common definition has been Gleason ≥ 3+4 and this was the primary definition which was used in IP1-PROSTAGRAM and for the majority of outcomes in this thesis.

There is growing evidence that this definition may be too stringent[177]. There is evidence that survival for men with certain types of Gleason ≥ 3+4 is similar to Gleason 3+3[392, 393]. In order to optimise screening tests for prostate cancer, it will be important that more accurate definitions of clinically significant prostate cancer are developed which account for factors such as the additional morphological subtypes of Gleason pattern 4 and tumour volume.

Long-term follow-up of men diagnosed with cancer in the IP1-PROSTAGRAM cohort could provide valuable information on the natural history of MRI screen-detected cancers. None of the existing definitions of significant disease consider the MRI visibility as a prognostic factor despite evidence that MRI visible cancer may be more aggressive than cancers not detected by MRI[394]. This is supported by basic science studies showing that MRI visible cancers may have aggressive genomic, transcriptomic, and pathological hallmarks which make them clinically more aggressive than non-visible MRI cancer[395]. In summary, it is clear that longitudinal studies of the IP1-PROSTAGRAM cohort are required and such studies have already been planned by linking the participants to NHS Digital's data repository.

12.5.3 Development of a scoring system for MRI screening

If the role of fast MRI is to be advanced for screening, there will need to be further work developing equivalent reporting and educational strategies as have been incorporated into prostate mpMRI training and screening programmes for other cancers. Previous studies have highlighted that interobserver variability is related to radiological experience and that it improves with dedicated experience and training[396]. In breast cancer screening, there are numerous quality assurance programmes to reduce interobserver variability. All mammograms are reported by certified radiologists and certification is based on specific criteria which include training, number of scans read per year and experience[397].

In addition self-assessment schemes such as the PERFORMS (Personal Performance in Mammographic Screening) in the UK provides radiologists with opportunities to improve their own performance[398]. Similar evidence-based programmes are required for prostate MRI screening given the findings in Chapter 6 that interobserver variability of the PI-RADS scoring system for screening was fair and has not reached the level achieved with mammograms for breast cancer screening using BI-RADS[399].

Equivocal MRI lesions are perhaps the most difficult assessment category with Chapter 6 showing considerable interobserver variability in the assessment of PI-RADS 3 (equivocal) lesions (kappa 0.211, AC1 0.466 and 67.9% agreement). Further work is required to develop

a formal scoring system for MRI lesions in a screening setting. The PI-RADS criteria were designed for men with a raised PSA and certain characterisations such as diffuse findings may require different scoring in a screening setting.

12.5.4 Optimising the re-screening interval for MRI

A key problem with PSA screening has been the recommendation for annual or biannual screening. With each screening round there is cumulative risk of false positives and overdiagnosis. The optimal re-screening interval for an MRI-based screening programme needs further research given the prolonged natural history of screen-detected prostate cancer. If a screen-negative MRI confers significant reassurance against future risk of cancer, the MRI screening pathway might be delivered at longer intervals and with fewer screening rounds. Interval studies in this area are clearly warranted in order to investigate this.

It is possible that the high accuracy of fast MRI could permit a safe extension of screening intervals. Other tumours groups have shown that the introduction of a more sensitive test provides an opportunity for extended screening intervals[400]. For example, in cervical cancer the shift from primary Papanicolaou cytologic testing to primary human papillomavirus (HPV) screening may allow retesting intervals to be extended beyond three years[401] to at least five year intervals[402].

12.6 A framework for future research

Screening tests can have a large impact on public health and careful evaluation is required prior to their introduction. While this thesis has provided the initial evidence for fast MRI, there needs to be much more work to reproduce and expand on these findings. It is important to consider a range of alternative methods to incorporate MRI into screening given that even a short, non-contrast MRI protocol has considerable additional costs compared to PSA. These will need to be carefully evaluated in future studies and in isolation this thesis cannot be considered as providing sufficient evidence to change screening practice.

There are useful lessons to be learnt from the widespread introduction of opportunistic PSA screening. PSA was developed in the 1980s for prostate cancer surveillance rather than as a diagnostic or screening test[45]. Its widespread adoption for screening began in the 1990s after paired cohort studies showed that PSA had better diagnostic performance than DRE[48]. The prolonged natural history of prostate cancer means that it can take over a decade for RCT

outcomes to mature, and widespread adoption of PSA occurred prior to RCTs' evidence which highlighted the uncertain balance between the benefits and harms of PSA screening.

There are significant risks to population health from rapid implementation of a new screening test prior to evidence from RCTs. Even if RCTs suggest that the harms outweigh the risks, once a screening test has acceptance by the public and the medical profession, it can be challenging to reduce demand for the test. Previous attempts to downgrade PSA screening in international guidelines have drawn significant criticism from public and professionals[403]. In addition, for researchers evaluating screening tests there needs to be sufficient numbers of men who will not be exposed to the test in order to ensure reliable results in clinical trials. This is not possible in countries such as the USA where in 2001 a population-based survey found that 75% men over 50 years were having PSA screening[49]. This has contaminated the control arms of RCTs such as PLCO and the ambiguity in the evidence has created widespread debate on the role of prostate cancer screening.

To avoid similar issues, this thesis concludes by strongly recommending that the evaluation of fast MRI follows a stepwise framework to ensure its reliable evaluation as a screening test. Unlike the phases of drug development or medical devices there has been a less well-defined pathway for evaluating screening tests. A specific framework has recently been developed for diagnostic and screening tests, known as the CanTest Framework[404]. This recognises the specific challenges for evaluating a cancer screening test and accounts for issues such as the low prevalence of disease, agreeing a suitable reference standard and other factors such as overdiagnosis.

In this framework, implementation of MRI screening can be conceptualised as progressing through five consecutive phases, as shown in Figure 84;

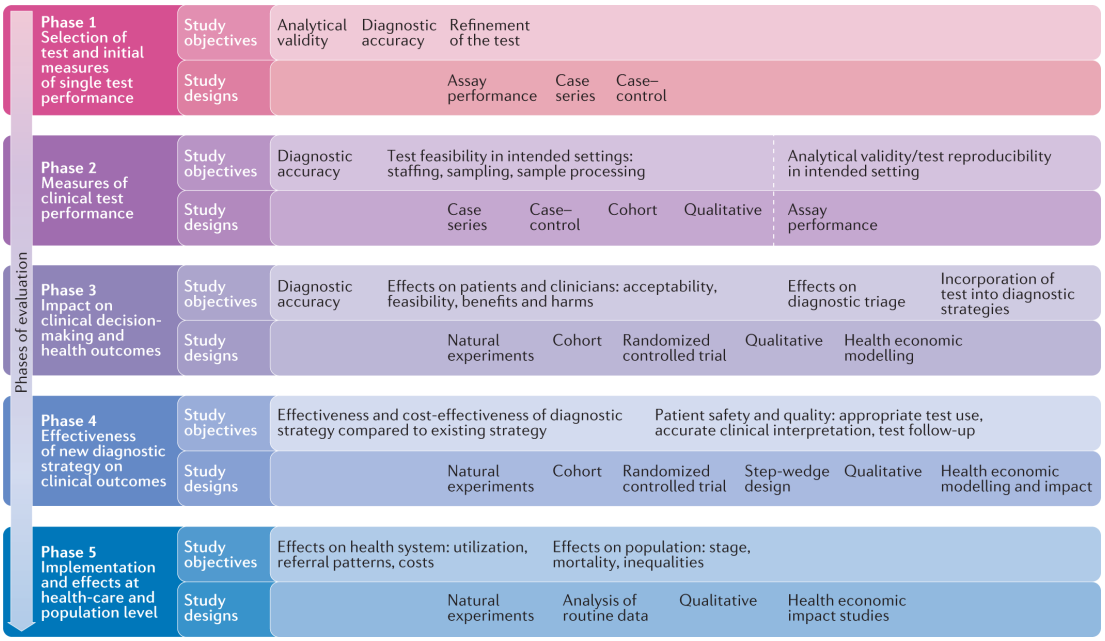


Figure 81: Research framework for implementing fast MRI as a screening test.

- Phase 1 involves the technical development of an MRI screening protocol and conducting case-series or pilot studies as proof of principle. This was completed in Chapter 4 of this study and a pilot study of MRI was completed by Nam et al[119].

- Phase 2 involves completing a number of observational studies which will evaluate either feasibility[405] or baseline performance measures of MRI in a screening setting[406]. This phase was completed in Chapters 6-8.

- Phase 3: Following this thesis we have entered Phase 3 which requires combining different trial designs to estimate the impacts on health outcomes as described in the preceding section. Such studies provide an indication on the impact on mortality but cannot be accepted as proof of efficacy of MRI screening as they are prone to lead-time, length-time and overdiagnosis bias. The Imperial Prostate Research group has been applying for funding for a Phase 3 randomised controlled trial evaluating diagnostic utility of fast MRI, drawing heavily on the experience and results described in this thesis.

- Phase 4: If Phase 3 trials are successful, this stage requires a prospective large-scale, randomised controlled screening trial to establish the effect of MRI screening on prostate cancer mortality. This would follow a similar design and end-point of mortality reduction as has been completed in mammography for breast cancer[407], HPV-based screening for cervical cancer[408] and stool-based tests for bowel cancer[409]. For prostate cancer, the primary end-point would be to compare prostate-cancer

specific mortality in MRI pathway group versus a control group of incidental PSA testing with shared decision making as recommended in most healthcare settings.

We have not yet reached a stage where the evidence supports a funding application for such a large-scale clinical trial which would need more than a decade to accrue sufficient events to provide results. The PLCO and ERSPC studies have run for many years and screened between 75,000 to 160,000 men. Completing such large-scale randomised trials is a substantial logistical and technical challenge so can only be considered at a late stage once the evidence-base is sufficiently robust. Nevertheless once these studies are completed the results can be used to perform a robust cost-effectiveness analysis allowing the final decision to be made as to whether fast MRI is viable to implement at a population level.

- Phase 5 is the post-implementation phase and is a standard step when evaluating the impact of any public health measure. This phase is usually conducted by public health experts and health economists. It involves a series of observational studies evaluating various performance metrics of the new programme and impact on population health.

12.7 Final Remarks

The UK National Screening Committee has been calling for further research into alternative screening tests for prostate cancer. This thesis has proposed MRI as an alternative test and argued that it may have certain characteristics which make it attractive for screening. Evidence was found for a fast MRI protocol with a higher threshold for denoting a screen-positive result. Rather than as an independent test, fast MRI may have optimal performance within a multi-modal pathway with a serum biomarker. Despite these encouraging findings, there are numerous challenges and complexities associated with MRI: a high interobserver variability, management of indeterminate lesions and limited MRI capacity.

Nevertheless, with continued developments in MRI technology, the future appears promising for the application of MRI in prostate cancer screening. The process of evaluating MRI as a screening test is now underway and lessons must be learnt from the challenges faced in PSA screening. It is imperative that the phased stepwise research approach set out in this thesis is followed before MRI-based screening is even considered for implementation in the general population.

Bibliography

1. UK National Screening Committee. The UK NSC recommendation on Prostate cancer screening and PSA testing in men over the age of 50 2016 [Available from: https://legacyscreening.phe.org.uk/prostatecancer.

2. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer; 2018. 2018.

3. Carioli G, Bertuccio P, Boffetta P, Levi F, La Vecchia C, Negri E, et al. European cancer mortality predictions for the year 2020 with a focus on prostate cancer. Annals of Oncology. 2020;31(5):650-8.

4. Lloyd T, Hounsome L, Mehay A, Mee S, Verne J, Cooper A. Lifetime risk of being diagnosed with, or dying from, prostate cancer by major ethnic group in England 2008–2010. BMC Medicine. 2015;13:171.

5. Nations U. World Population Ageing 2015 (ST/ESA/SER. A/390). Department of Economic and Social Affairs, Population Division New York, NY; 2015.

6. World Health Organization. Cancer tomorrow. Estimated number of deaths from 2018 to 2040, prostate, males, all ages: International Agency for Research on Cancer; [Available from: https://gco.iarc.fr/tomorrow/home.

7. Yin M, Bastacky S, Chandran U, Becich MJ, Dhir R. Prevalence of incidental prostate cancer in the general population: a study of healthy organ donors. The Journal of urology. 2008;179(3):892-5; discussion 5.

8. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the United Kingdom, 2010–2040. British Journal of Cancer. 2012;107(7):1195-202.

9. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, et al. The case for early detection. Nature Reviews Cancer. 2003;3(4):243-52.

10. Wilson JMG, Jungner G, Organization WH. Principles and practice of screening for disease. 1968.

11. Cancer Research UK. Cancer Mortality Statistics 2017 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality.

12. World Health Organization. Deaths(#), males, Malignant neoplasm of prostate 2017 [9 July 2020]. Available from: https://gateway.euro.who.int/en/indicators/hfamdb_575-deaths-males-malignant-neoplasm-of-prostate/.

13. Culp MB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. European Urology. 2020;77(1):38-52.

14. Bostwick DG, Burke HB, Djakiew D, Euling S, Ho Sm, Landolph J, et al. Human prostate cancer risk factors. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2004;101(S10):2371-490.

15. Office for National Statistics. Deaths from prostate cancer, England and Wales, 2001 to 2017 2019 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/009669deathsfromprostatecancerenglandandwales2001to2017.

16. Feuer EJ, Merrill RM, Hankey BF. Cancer Surveillance Series: Interpreting Trends in Prostate Cancer Part II: Cause of Death Misclassification and the Recent Rise and Fall in Prostate Cancer Mortality. JNCI Journal of the National Cancer Institute. 1999;91(12):1025-32.

17. Bell KJL, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. International Journal of Cancer. 2015;137(7):1749-57.

18. Committee UNS. The UK NSC recommendation on prostate cancer screening/PSA testing in men over the age of 50. 2018.

19. Ahmed HU, Arya M, Freeman A, Emberton M. Do low-grade and low-volume prostate cancers bear the hallmarks of malignancy? The lancet oncology. 2012;13(11):e509-e17.

20. Stamey TA, Freiha FS, McNeal JE, Redwine EA, Whittemore AS, Schmid HP. Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. Cancer. 1993;71(S3):933-8.

21. Winkler MH, Livni N, Mannion EM, Hrouda D, Christmas T. Characteristics of incidental prostatic adenocarcinoma in contemporary radical cystoprostatectomy specimens. BJU International. 2007;99(3):554-8.

22. Wolters T, Roobol MJ, van Leeuwen PJ, van den Bergh RCN, Hoedemaeker RF, van Leenders GJLH, et al. A Critical Analysis of the Tumor Volume Threshold for Clinically Insignificant Prostate Cancer Using a Data Set of a Randomized Screening Trial. The Journal of Urology. 2011;185(1):121-5.

23. Schoots IG, van der Kwast TH. MR Imaging in Prostate Tumor Volume Assessment: How Accurate? In: Klotz L, editor. Active Surveillance for Localized Prostate Cancer: A New Paradigm for Clinical Management. Cham: Springer International Publishing; 2018. p. 95-109.

24. Stamey TA, McNeal JE, Yemoto CM, Sigal BM, Johnstone IM. Biological determinants of cancer progression in men with prostate cancer. Jama. 1999;281(15):1395-400.

25. Eichelberger LE, Koch MO, Eble JN, Ulbright TM, Juliar BE, Cheng L. Maximum tumor diameter is an independent predictor of prostate-specific antigen recurrence in prostate cancer. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2005;18(7):886-90.

26. Carvalhal GF, Humphrey PA, Thorson P, Yan Y, Ramos CG, Catalona WJ. Visual estimate of the percentage of carcinoma is an independent predictor of prostate carcinoma recurrence after radical prostatectomy. Cancer. 2000;89(6):1308-14.

27. van Oort IM, Witjes JA, Kok DE, Kiemeney LA, Hulsbergen-vandeKaa CA. Maximum tumor diameter is not an independent prognostic factor in high-risk localized prostate cancer. World J Urol. 2008;26(3):237-41.

28. Ohori M, Wheeler TM, Kattan MW, Goto Y, Scardino PT. Prognostic significance of positive surgical margins in radical prostatectomy specimens. J Urol. 1995;154(5):1818-24.

29. Salomon L, Levrel O, Anastasiadis AG, Irani J, De La Taille A, Saint F, et al. Prognostic Significance of Tumor Volume after Radical Prostatectomy: A Multivariate Analysis of Pathological Prognostic Factors. European Urology. 2003;43(1):39-44.

30. Epstein JI, Carmichael M, Partin AW, Walsh PC. Is tumor volume an independent predictor of progression following radical prostatectomy? A multivariate analysis of 185 clinical stage B adenocarcinomas of the prostate with 5 years of followup. J Urol. 1993;149(6):1478-81.

31. Kikuchi E, Scardino PT, Wheeler TM, Slawin KM, Ohori M. Is tumor volume an independent prognostic factor in clinically localized prostate cancer? The Journal of urology. 2004;172(2):508-11.

32. Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. BJU international. 2013;111(5):753-60.

33. Mohler JL, Antonarakis ES, Armstrong AJ, D'Amico AV, Davis BJ, Dorff T, et al. Prostate cancer, version 2.2019, NCCN clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network. 2019;17(5):479-505.

34. Ahmed HU, Hu Y, Carter T, Arumainayagam N, Lecornet E, Freeman A, et al. Characterizing Clinically Significant Prostate Cancer Using Template Prostate Mapping Biopsy. 2011;186(2):458-64.

35. NICE Guidance - Prostate cancer: diagnosis and management. BJU international. 2019;124(1):9-26.

36. Chen RC, Rumble RB, Loblaw DA, Finelli A, Ehdaie B, Cooperberg MR, et al. Active Surveillance for the Management of Localized Prostate Cancer (Cancer Care Ontario Guideline): American Society of Clinical Oncology Clinical Practice Guideline Endorsement. 2016.

37. Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. New England Journal of Medicine. 2016;375(15):1415-24.

38. Bill-Axelson A, Holmberg L, Garmo H, Taari K, Busch C, Nordling S, et al. Radical Prostatectomy or Watchful Waiting in Prostate Cancer — 29-Year Follow-up. New England Journal of Medicine. 2018;379(24):2319-29.

39. Epstein JI, Walsh PC, Carmichael M, Brendler CB. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage t1 c) prostate cancer. Jama. 1994;271(5):368-74.

40. Epstein JI, Allsbrook Jr WC, Amin MB, Egevad LL, Committee IG. The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. The American journal of surgical pathology. 2005;29(9):1228-42.

41. Ahmed HU, Hu Y, Carter T, Arumainayagam N, Lecornet E, Freeman A, et al. Characterizing clinically significant prostate cancer using template prostate mapping biopsy. The Journal of urology. 2011;186(2):458-64.

42. Grönberg H, Adolfsson J, Aly M, Nordström T, Wiklund P, Brandberg Y, et al. Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. The Lancet Oncology. 2015;16(16):1667-76.

43. Verbeek JFM, Bangma CH, Kweldam CF, Van Der Kwast TH, Kümmerlin IP, Van Leenders GJLH, et al. Reducing unnecessary biopsies while detecting clinically significant prostate cancer including cribriform growth with the ERSPC Rotterdam risk calculator and 4Kscore. Urologic Oncology: Seminars and Original Investigations. 2019;37(2):138-44.

44. Stamey TA. Biological Determinants of Cancer Progression in Men With Prostate Cancer. JAMA. 1999;281(15):1395.

45. Catalona WJ. History of the discovery and clinical translation of prostate-specific antigen. 2014;1(1):12-4.

46. Thompson IM, Tangen CM, Goodman PJ, Lucia MS, Parnes HL, Lippman SM, et al. Finasteride Improves the Sensitivity of Digital Rectal Examination for Prostate Cancer Detection. Journal of Urology. 2007;177(5):1749-52.

47. Chodak GW, Keller P, Schoenberg HW. Assessment of Screening for Prostate Cancer Using the Digital Rectal Examination. Journal of Urology. 1989;141(5):1136-8.

48. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJJ, et al. Measurement of Prostate-Specific Antigen in Serum as a Screening Test for Prostate Cancer. New England Journal of Medicine. 1991;324(17):1156-61.

49. Sirovich BE, Schwartz LM, Woloshin S. Screening Men for Prostate and Colorectal Cancer in the United States. JAMA. 2003;289(11):1414.

50. Jemal A, Center MM, DeSantis C, Ward EM. Global Patterns of Cancer Incidence and Mortality Rates and Trends. Cancer Epidemiology Biomarkers & Prevention. 2010;19(8):1893-907.

51. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer. 2015;136(5):E359-E86.

52. Office for National Statistics. Cancer registration statistics 2017 [Available from: https://www.ons.gov.uk/releases/cancerregistrationstatisticsengland2017.

53. Patel AR, Klein EA. Risk factors for prostate cancer. Nature Clinical Practice Urology. 2009;6(2):87-95.

54. Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, et al. The Influence of Finasteride on the Development of Prostate Cancer. New England Journal of Medicine. 2003;349(3):215-24.

55. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. The Cochrane database of systematic reviews. 2013(1):Cd004720.

56. Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V, et al. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. European Urology. 2019;76(1):43-51.

57. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M, et al. The benefits and harms of breast cancer screening: an independent review: A report jointly commissioned by Cancer Research UK and the Department of Health (England) October 2012. British Journal of Cancer. 2013;108(11):2205-40.

58. Wu GH-M, Auvinen A, Yen AM-F, Hakama M, Tammela TL, Stenman U-H, et al. The Impact of Interscreening Interval and Age on Prostate Cancer Screening With Prostate-Specific Antigen. European Urology. 2012;61(5):1011-8.

59. Carroll PH, Mohler JL. NCCN guidelines updates: prostate cancer and prostate cancer early detection. Journal of the National Comprehensive Cancer Network. 2018;16(5S):620-3.

60. Force UPST. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation StatementUSPSTF Recommendation: Screening for Prostate CancerUSPSTF Recommendation: Screening for Prostate Cancer. Jama. 2018;319(18):1901-13.

61. Gandaglia G, Albers P, Abrahamsson P-A, Briganti A, Catto JWF, Chapple CR, et al. Structured Population-based Prostate-specific Antigen Screening for Prostate Cancer: The European Association of Urology Position in 2019. European Urology. 2019;76(2):142-50.

62. Pinsky PF, Prorok PC, Yu K, Kramer BS, Black A, Gohagan JK, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. Cancer. 2017;123(4):592-9.

63. Gulati R, Tsodikov A, Wever EM, Mariotto AB, Heijnsdijk EAM, Katcher J, et al. The impact of PLCO control arm contamination on perceived PSA screening efficacy. Cancer Causes & Control. 2012;23(6):827-35.

64. Schröder FH, Roobol MJ. ERSPC and PLCO Prostate Cancer Screening Studies: What Are the Differences? 2010;58(1):46-52.

65. Tsodikov A, Gulati R, Heijnsdijk EAM, Pinsky PF, Moss SM, Qiu S, et al. Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials. Annals of Internal Medicine. 2017;167(7):449.

66. De Koning HJ, Gulati R, Moss SM, Hugosson J, Pinsky PF, Berg CD, et al. The efficacy of prostate-specific antigen screening: Impact of key components in the ERSPC and PLCO trials. Cancer. 2018;124(6):1197-206.

67. Martin RM, Donovan JL, Turner EL, Metcalfe C, Young GJ, Walsh EI, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality. Jama. 2018;319(9):883.

68. Lundgren P-O, Kjellman A, Norming U, Gustafsson O. Long-Term Outcome of a Single Intervention Population Based Prostate Cancer Screening Study. Journal of Urology. 2018;200(1):82-8.

69. Kilpeläinen TP, Tammela TLJ, Roobol M, Hugosson J, Ciatto S, Nelen V, et al. False-positive screening results in the European randomized study of screening for prostate cancer. European Journal of Cancer. 2011;47(18):2698-705.

70. Sandblom G, Varenhorst E, Löfman O, Rosell J, Carlsson P. Clinical Consequences of Screening for Prostate Cancer: 15 Years Follow-up of a Randomised Controlled Trial in Sweden. European Urology. 2004;46(6):717-24.

71. Sandblom G, Varenhorst E, Rosell J, Lofman O, Carlsson P. Randomised prostate cancer screening trial: 20 year follow-up. BMJ. 2011;342(mar31 1):d1539-d.

72. Labrie F, Candas B, Cusan L, Gomez JL, Bélanger A, Brousseau G, et al. Screening decreases prostate cancer mortality: 11-year follow-up of the 1988 Quebec prospective randomized controlled trial. The Prostate. 2004;59(3):311-8.

73. Ilic D, Djulbegovic M, Jung JH, Hwang EC, Zhou Q, Cleves A, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. BMJ. 2018:k3519.

74. Loeb S, Bjurlin M, Nicholson J, Tammela TL, Penson D, Carter HB, et al. Overdiagnosis and Overtreatment of Prostate Cancer. European urology. 2014;65(6):1046-55.

75. Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schroder FH, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the

European Randomized Study of Screening for Prostate Cancer. Journal of the National Cancer Institute. 2003;95(12):868-78.

76. Quon H, Loblaw A, Nam R. Dramatic increase in prostate cancer cases by 2021. BJU international. 2011;108(11):1734-8.

77. Donovan JL, Hamdy FC, Lane JA, Mason M, Metcalfe C, Walsh E, et al. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. New England Journal of Medicine. 2016;375(15):1425-37.

78. Heijnsdijk EAM, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, et al. Quality-of-life effects of prostate-specific antigen screening. The New England journal of medicine. 2012;367(7):595-605.

79. Wilt TJ, Vo TN, Langsetmo L, Dahm P, Wheeler T, Aronson WJ, et al. Radical Prostatectomy or Observation for Clinically Localized Prostate Cancer: Extended Follow-up of the Prostate Cancer Intervention Versus Observation Trial (PIVOT). European Urology. 2020;77(6):713-24.

80. National Prostate Cancer Audit. Annual Report 2019 [Available from: https://www.npca.org.uk/content/uploads/2020/01/NPCA-Annual-Report-2019_090120.pdf.

81. Loeb S, Folkvaljon Y, Curnyn C, Robinson D, Bratt O, Stattin P. Uptake of Active Surveillance for Very-Low-Risk Prostate Cancer in Sweden. JAMA Oncology. 2017;3(10):1393.

82. Sampurno F, Earnest A, Kumari PB, Millar JL, Davis ID, Murphy DG, et al. Quality of care achievements of the Prostate Cancer Outcomes Registry–Victoria. Medical Journal of Australia. 2016;204(8):319-.

83. Guy D, Ghanem G, Loblaw A, Buckley R, Persaud B, Cheung P, et al. Diagnosis, referral, and primary treatment decisions in newly diagnosed prostate cancer patients in a multidisciplinary diagnostic assessment program. 2016;10(3-4):120.

84. Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA. 2015;314(1):80.

85. Kinsella N, Stattin P, Cahill D, Brown C, Bill-Axelson A, Bratt O, et al. Factors Influencing Men's Choice of and Adherence to Active Surveillance for Low-risk Prostate Cancer: A Mixed-method Systematic Review. European Urology. 2018;74(3):261-80.

86. Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. Jama. 2005;294(1):66-70.

87. Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM. Effect of Verification Bias on Screening for Prostate Cancer by Measurement of Prostate-Specific Antigen. New England Journal of Medicine. 2003;349(4):335-42.

88. Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. The Lancet. 2017;389(10071):815-22.

89. King CR, McNeal JE, Gill H, Presti JC, Jr. Extended prostate biopsy scheme improves reliability of Gleason grading: implications for radiotherapy patients. International journal of radiation oncology, biology, physics. 2004;59(2):386-91.

90. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. European urology. 2017;71(4):618-29.

91. Roddam AW, Hamdy FC, Allen NE, Price CP. The impact of reducing the prostate-specific antigen threshold and including isoform reflex tests on the performance characteristics of a prostate-cancer detection programme. BJU international. 2007;100(3):514-7.

92. Holmström B, Johansson M, Bergh A, Stenman U-H, Hallmans G, Stattin P. Prostate specific antigen for early detection of prostate cancer: longitudinal study. BMJ. 2009;339.

93. Wright EJ, Fang J, Metter EJ, Partin AW, Landis P, Chan DW, et al. Prostate specific antigen predicts the long-term risk of prostate enlargement: results from the Baltimore Longitudinal Study of Aging. The Journal of urology. 2002;167(6):2484-7.

94. Nam RK, Saskin R, Lee Y, Liu Y, Law C, Klotz LH, et al. Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy. The Journal of urology. 2010;183(3):963-8.

95. Raaijmakers R, Kirkels WJ, Roobol MJ, Wildhagen MF, Schrder FH. Complication rates and risk factors of 5802 transrectal ultrasound-guided sextant biopsies of the prostate within a population-based screening program. Urology. 2002;60(5):826-30.

96. Nagler HM, Gerber EW, Homel P, Wagner JR, Norton J, Lebovitch S, et al. Digital rectal examination is barrier to population-based prostate cancer screening. Urology. 2005;65(6):1137-40.

97. Hoffman RM, Lewis CL, Pignone MP, Couper MP, Barry MJ, Elmore JG, et al. Decision-Making Processes for Breast, Colorectal, and Prostate Cancer Screening: The DECISIONS Survey. Medical Decision Making. 2010;30(5_suppl):53-64.

98. Volk RJ, Hawley ST, Kneuper S, Holden EW, Stroud LA, Cooper CP, et al. Trials of Decision Aids for Prostate Cancer Screening. 2007;33(5):428-34.e11.

99. Evans R, Edwards A, Brett J, Bradburn M, Watson E, Austoker J, et al. Reduction in uptake of PSA tests following decision aids: systematic review of current aids and their evaluations. 2005;58(1):13-26.

100. Burns R, Walsh B, O'Neill S, O'Neill C. An examination of variations in the uptake of prostate cancer screening within and between the countries of the EU-27. Health Policy. 2012;108(2-3):268-76.

101. Melia J, Moss S, Johns L. Rates of prostate-specific antigen testing in general practice in England and Wales in asymptomatic and symptomatic patients: a cross-sectional study. BJU international. 2004;94(1):51-6.

102. Williams N, Hughes LJ, Turner EL, Donovan JL, Hamdy FC, Neal DE, et al. Prostate-specific antigen testing rates remain low in UK general practice: a cross-sectional study in six English cities. BJU international. 2011;108(9):1402-8.

103. Walsh B, Silles M, O'Neill C. The importance of socio-economic variables in cancer screening participation: A comparison between population-based and opportunistic screening in the EU-15. Health Policy. 2011;101(3):269-76.

104. Bell KJ, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. International journal of cancer. 2015;137(7):1749-57.

105. Wolf AM, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA: a cancer journal for clinicians. 2010;60(2):70-98.

106. Carroll PR, Parsons JK, Andriole G, Bahnson RR, Barocas DA, Castle EP, et al. Prostate cancer early detection, version 2.2015. Journal of the National Comprehensive Cancer Network. 2015;13(12):1534-61.

107. Carter HB, Albertsen PC, Barry MJ, Etzioni R, Freedland SJ, Greene KL, et al. Early Detection of Prostate Cancer: AUA Guideline. The Journal of Urology. 2013;190(2):419-26.

108. Heidenreich A, Abrahamsson P-A, Artibani W, Catto J, Montorsi F, Van Poppel H, et al. Early Detection of Prostate Cancer: European Association of Urology Recommendation. European Urology. 2013;64(3):347-54.

109. Louie K. UKNSC Screening for Prostate Cancer Review 2015 update. UK National Screening Committee. 2015.

110. Taskforce PC. Diagnosis and Management of Prostate Cancer in New Zealand Men: Recommendations from the Prostate Cancer Taskforce. . 2012.

111. Australia Prostate CFoAaCC. PSA Testing and Early Management of Test-Detected Prostate Cancer. 2016.

112. Bell N, Gorber SC, Shane A, Joffres M, Singh H, Dickinson J, et al. Recommendations on screening for prostate cancer with the prostate-specific antigen test. Canadian Medical Association Journal. 2014;186(16):1225-34.

113. Ávila M, Patel L, López S, Cortés-Sanabria L, Garin O, Pont À, et al. Patient-reported outcomes after treatment for clinically localized prostate cancer: A systematic review and meta-analysis. Cancer Treatment Reviews. 2018;66:23-44.

114. Yap T, Ahmed HU, Hindley RG, Guillaumier S, McCartan N, Dickinson L, et al. The Effects of Focal Therapy for Prostate Cancer on Sexual Function: A Combined Analysis of Three Prospective Trials. European Urology. 2016;69(5):844-51.

115. Johansson E, Steineck G, Holmberg L, Johansson J-E, Nyberg T, Ruutu M, et al. Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial. The Lancet Oncology. 2011;12(9):891-9.

116. . accessed 15/06/2017: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme. p. United Kingdom National Screening Committee.

117. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJY, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015;64(10):1637-49.

118. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2017: A review of current American Cancer Society guidelines and current issues in cancer screening. CA: a cancer journal for clinicians. 2017;67(2):100-21.

119. Nam RK, Wallis CJD, Stojcic-Bendavid J, Milot L, Sherman C, Sugar L, et al. A Pilot Study to Evaluate the Role of Magnetic Resonance Imaging for Prostate Cancer Screening in the General Population. The Journal of Urology. 2016;196(2):361-6.

120. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. Social Science & Medicine. 2005;61(2):417-30.

121. Alberts A, Schoots I, Drost FJ, Bokhorst L, Van Leenders G, Dwarkasing R, et al. Value of magnetic resonance imaging in population-based prostate cancer screening: Comparison of 3 biopsy strategies in the 5th screening round of the ERSPC Rotterdam. European Urology Supplements. 2017;16(3):e867-e8.

122. Alberts AR, Schoots IG, Bokhorst LP, Drost F-JH, Van Leenders GJ, Krestin GP, et al. Characteristics of Prostate Cancer Found at Fifth Screening in the European Randomized Study of Screening for Prostate Cancer Rotterdam: Can We Selectively Detect High-grade Prostate Cancer with Upfront Multivariable Risk Stratification and Magnetic Reson. European Urology. 2018;73(3):343-50.

123. Boesen L, Nørgaard N, Løgager V, Balslev I, Bisbjerg R, Thestrup K-C, et al. Assessment of the Diagnostic Accuracy of Biparametric Magnetic Resonance Imaging for Prostate Cancer in Biopsy-Naive Men. JAMA Network Open. 2018;1(2):e180219.

124. de Bekker-Grob E, Rose J, Donkers B, Essink-Bot M-L, Bangma C, Steyerberg E. Men's preferences for prostate cancer screening: a discrete choice experiment. British journal of cancer. 2013;108(3):533-41.

125. Chen Z, Zheng Y, Ji G, Liu X, Li P, Cai L, et al. Accuracy of dynamic contrast-enhanced magnetic resonance imaging in the diagnosis of prostate cancer: systematic review and meta-analysis. Oncotarget. 2017;8(44).

126. de Rooij M, Crienen S, Witjes JA, Barentsz JO, Rovers MM, Grutters JPC. Cost-effectiveness of Magnetic Resonance (MR) Imaging and MR-guided Targeted Biopsy Versus Systematic Transrectal Ultrasound–Guided Biopsy in Diagnosing Prostate Cancer: A Modelling Study from a Health Care Perspective. European Urology. 2014;66(3):430-6.

127. De Visschere PJL, Naesens L, Libbrecht L, Van Praet C, Lumen N, Fonteyne V, et al. What kind of prostate cancers do we miss on multiparametric magnetic resonance imaging? European Radiology. 2016;26(4):1098-107.

128. Drost FJH, Osses DF, Nieboer D, Steyerberg EW, Bangma CH, Roobol MJ, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. Cochrane Database of Systematic Reviews. 2019(4).

129. Fütterer JJ, Briganti A, De Visschere P, Emberton M, Giannarini G, Kirkham A, et al. Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature. European Urology. 2015;68(6):1045-53.

130. Hamoen EH, de Rooij M, Witjes JA, Barentsz JO, Rovers MM. Use of the prostate imaging reporting and data system (PI-RADS) for prostate cancer detection with multiparametric magnetic resonance imaging: a diagnostic meta-analysis. European urology. 2015;67(6):1112-21.

131. Jambor I, Verho J, Ettala O, Knaapila J, Taimen P, Syvänen KT, et al. Validation of IMPROD biparametric MRI in men with clinically suspected prostate cancer: A prospective multi-institutional trial. PLOS Medicine. 2019;16(6):e1002813.

132. Kang Z, Min X, Weinreb J, Li Q, Feng Z, Wang L. Abbreviated biparametric versus standard multiparametric MRI for diagnosis of prostate cancer: a systematic review and meta-analysis. American Journal of Roentgenology. 2019;212(2):357-65.

133. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. New England Journal of Medicine. 2018;378(19):1767-77.

134. Miah S, Hosking-Jervis F, Connor MJ, Eldred-Evans D, Shah TT, Arya M, et al. A Multicentre Analysis of the Detection of Clinically Significant Prostate Cancer Following Transperineal Image-fusion Targeted and Nontargeted Systematic Prostate Biopsy in Men at Risk. European Urology Oncology. 2019.

135. Moore CM, Robertson NL, Arsanious N, Middleton T, Villers A, Klotz L, et al. Image-guided prostate biopsy using magnetic resonance imaging-derived targets: a systematic review. European urology. 2013;63(1):125-40.

136. Neves J, Simmons L, Kanthabalan A, McCartan N, Freeman A, Punwani S, et al. Combining mpMRI sequences for the diagnosis of prostate cancer – the value of adding diffusion and contrast enhancement to T2W on 3Tesla: Outcomes from the PICTURE trial. European Urology Supplements.15(12):e1505-e6.

137. Niu X-k, Chen X-h, Chen Z-f, Chen L, Li J, Peng T. Diagnostic performance of biparametric MRI for detection of prostate cancer: a systematic review and meta-analysis. American Journal of Roentgenology. 2018;211(2):369-78.

138. Rastinehad AR, Waingankar N, Turkbey B, Yaskiv O, Sonstegard AM, Fakhoury M, et al. Comparison of multiparametric MRI scoring systems and the impact on cancer detection in patients undergoing MR US fusion guided prostate biopsies. PloS one. 2015;10(11):e0143404.

139. Rosenkrantz AB, Lepor H, Huang WC, Taneja SS. Practical Barriers to Obtaining Pre-Biopsy Prostate MRI: Assessment in Over 1,500 Consecutive Men Undergoing Prostate Biopsy in a Single Urologic Practice. Urologia Internationalis. 2016;97(2):247-8.

140. Rouvière O, Puech P, Renard-Penna R, Claudon M, Roy C, Mège-Lechevallier F, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. The Lancet Oncology. 2019;20(1):100-9.

141. Schouten MG, van der Leest M, Pokorny M, Hoogenboom M, Barentsz JO, Thompson LC, et al. Why and Where do We Miss Significant Prostate Cancer with Multi-parametric Magnetic Resonance Imaging followed by Magnetic Resonance-guided and Transrectal Ultrasound-guided Biopsy in Biopsy-naïve Men? European Urology. 2017;71(6):896-903.

142. Stavrinides V, Giganti F, Orczyk C, Punwani S, Allen C, Kirkham A, et al. Visible disease at baseline accelerates time to exit from MRI-based active surveillance. European Urology Supplements. 2019;18(1):e607.

143. Venderink W, van Luijtelaar A, Bomers JG, van der Leest M, Hulsbergen-van de Kaa C, Barentsz JO, et al. Results of targeted biopsy in men with magnetic resonance imaging lesions classified equivocal, likely or highly likely to be clinically significant prostate cancer. European urology. 2018;73(3):353-60.

144. Weiss J, Martirosian P, Notohamiprodjo M, Kaufmann S, Othman AE, Grosse U, et al. Implementation of a 5-Minute Magnetic Resonance Imaging Screening Protocol for Prostate Cancer in Men With Elevated Prostate-Specific Antigen Before Biopsy. Investigative Radiology. 2018;53(3):186-90.

145. Wolters T, Roobol MJ, van Leeuwen PJ, van den Bergh RC, Hoedemaeker RF, van Leenders GJ, et al. A critical analysis of the tumor volume threshold for clinically insignificant

prostate cancer using a data set of a randomized screening trial. The Journal of urology. 2011;185(1):121-5.

146. Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Diagnostic performance of prostate imaging reporting and data system version 2 for detection of prostate cancer: a systematic review and diagnostic meta-analysis. European urology. 2017;72(2):177-88.

147. Woo S, Suh CH, Kim SY, Cho JY, Kim SH, Moon MH. Head-to-head comparison between biparametric and multiparametric MRI for the diagnosis of prostate cancer: a systematic review and meta-analysis. American Journal of Roentgenology. 2018:W226-W41.

148. Rastinehad AR, Waingankar N, Turkbey B, Yaskiv O, Sonstegard AM, Fakhoury M, et al. Comparison of Multiparametric MRI Scoring Systems and the Impact on Cancer Detection in Patients Undergoing MR US Fusion Guided Prostate Biopsies. PLOS ONE. 2015;10(11):e0143404.

149. Hamoen EHJ, De Rooij M, Witjes JA, Barentsz JO, Rovers MM. Use of the Prostate Imaging Reporting and Data System (PI-RADS) for Prostate Cancer Detection with Multiparametric Magnetic Resonance Imaging: A Diagnostic Meta-analysis. European Urology. 2015;67(6):1112-21.

150. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. Journal of clinical epidemiology. 2009;62(1):5-12.

151. Andriole GL, Levin DL, Crawford ED, Gelmann EP, Pinsky PF, Chia D, et al. Prostate Cancer Screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: Findings From the Initial Screening Round of a Randomized Trial. JNCI Journal of the National Cancer Institute. 2005;97(6):433-8.

152. Rouvière O, Puech P, Renard-Penna R, Claudon M, Roy C, Mège-Lechevallier F, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. The Lancet Oncology. 2019;20(1):100-9.

153. Van Der Marel J, Van Baars R, Rodriguez A, Quint WGV, Van De Sandt MM, Berkhof J, et al. The increased detection of cervical intraepithelial neoplasia when using a second biopsy at colposcopy. Gynecologic Oncology. 2014;135(2):201-7.

154. Lehman CD, Gatsonis C, Kuhl CK, Hendrick RE, Pisano ED, Hanna L, et al. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. The New England journal of medicine. 2007;356(13):1295-303.

155. Bernard JR, Vallow LA, Deperi ER, McNeil RB, Feigel DG, Amar S, et al. In Newly Diagnosed Breast Cancer, Screening MRI of the Contralateral Breast Detects Mammographically Occult Cancer, Even in Elderly Women: The Mayo Clinic in Florida Experience. The Breast Journal. 2010;16(2):118-26.

156. King TA, Gurevich I, Sakr R, Patil S, Stempel M, Morrow M. Occult malignancy in patients undergoing contralateral prophylactic mastectomy. Ann Surg. 2011;254(1):2-7.

157. Verbeek JFM, Roobol MJ. What is an acceptable false negative rate in the detection of prostate cancer? Translational Andrology and Urology. 2018;7(1):54-60.

158. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤4.0 ng per Milliliter. New England Journal of Medicine. 2004;350(22):2239-46.

159. Teo CH, Ng CJ, Booth A, White A. Barriers and facilitators to health screening in men: A systematic review. Social Science & Medicine. 2016;165:168-76.

160. Odedina FT, Akinremi TO, Chinegwundoh F, Roberts R, Yu D, Reams RR, et al. Prostate cancer disparities in Black men of African descent: a comparative literature review of prostate cancer burden among Black men in the United States, Caribbean, United Kingdom, and West Africa. Infectious Agents and Cancer. 2009;4(1):S2.

161. Austin KL, Power E, Solarin I, Atkin WS, Wardle J, Robb KA. Perceived barriers to flexible sigmoidoscopy screening for colorectal cancer among UK ethnic minority groups: a qualitative study. 2009;16(4):174-9.

162. Gluecker TM, Johnson CD, Harmsen WS, Offord KP, Harris AM, Wilson LA, et al. Colorectal Cancer Screening with CT Colonography, Colonoscopy, and Double-Contrast Barium Enema Examination: Prospective Assessment of Patient Perceptions and Preferences1. 2003;227(2):378-84.

163. Edwards JT, Mendelson RM, Fritschi L, Foster NM, Wood C, Murray D, et al. Colorectal Neoplasia Screening with CT Colonography in Average-Risk Asymptomatic Subjects: Community-based Study. Radiology. 2004;230(2):459-64.

164. Emberton M. Is Prostate Magnetic Resonance Imaging Going to Break the Bank? European Urology.66(3):437-8.

165. Grubb RL, Pinsky PF, Greenlee RT, Izmirlian G, Miller AB, Hickey TP, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. 2008;102(11):1524-30.

166. Otto SJ, Moss SM, Määttänen L, Roobol M, Zappa M, Nelen V, et al. PSA levels and cancer detection rate by centre in the European Randomized Study of Screening for Prostate Cancer. European Journal of Cancer. 2010;46(17):3053-60.

167. Chang DTS, Challacombe B, Lawrentschuk N. Transperineal biopsy of the prostate—is this the future? 2013;10(12):690-702.

168. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. European Urology. 2016;69(1):16-40.

169. Swensen SJ, Jett JR, Sloan JA, Midthun DE, Hartman TE, Sykes A-M, et al. Screening for lung cancer with low-dose spiral computed tomography. American journal of respiratory and critical care medicine. 2002;165(4):508-13.

170. Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, et al. American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography. CA: A Cancer Journal for Clinicians. 2007;57(2):75-89.

171. Venderink W, Van Luijtelaar A, Bomers JGR, Van Der Leest M, Hulsbergen-Van De Kaa C, Barentsz JO, et al. Results of Targeted Biopsy in Men with Magnetic Resonance Imaging Lesions Classified Equivocal, Likely or Highly Likely to Be Clinically Significant Prostate Cancer. European Urology. 2018;73(3):353-60.

172. Vargas HA, Hötker AM, Goldman DA, Moskowitz CS, Gondo T, Matsumoto K, et al. Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference. European Radiology. 2016;26(6):1606-12.

173. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. European urology. 2016;69(1):16-40.

174. Stamey TA, Freiha FS, McNeal JE, Redwine EA, Whittemore AS, Schmid H-P. Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. Cancer. 1993;71(S3):933-8.

175. Wolters T, Roobol MJ, van Leeuwen PJ, van den Bergh RCN, Hoedemaeker RF, van Leenders GJLH, et al. A critical analysis of the tumor volume threshold for clinically insignificant prostate cancer using a data set of a randomized screening trial. The Journal of urology. 2011;185(1):121-5.

176. Epstein JI, Allsbrook WC, Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. 2005;29(9):1228-42.

177. Matoso A, Epstein JI. Defining clinically significant prostate cancer on the basis of pathological findings. Histopathology. 2019;74(1):135-45.

178. Chin M, Mendelson R, Edwards J, Foster N, Forbes G. Computed tomographic colonography: prevalence, nature, and clinical significance of extracolonic findings in a community screening program. The American journal of gastroenterology. 2005;100(12):2771-6.

179. Morgan L, Choi H, Reid M, Khawaja A, Mazzone PJ. Frequency of Incidental Findings and Subsequent Evaluation in Low-Dose Computed Tomographic Scans for Lung Cancer Screening. Annals of the American Thoracic Society. 2017;14(9):1450-6.

180. Barentsz JO, Weinreb JC, Verma S, Thoeny HC, Tempany CM, Shtern F, et al. Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use. European urology. 2016;69(1):41-9.

181. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. European Urology. 2019;76(3):340-51.

182. Brizmohun Appayya M, Adshead J, Ahmed HU, Allen C, Bainbridge A, Barrett T, et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection - recommendations from a UK consensus meeting. BJU International. 2018;122(1):13-25.

183. Rosenkrantz AB, Ginocchio LA, Cornfeld D, Froemming AT, Gupta RT, Turkbey B, et al. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. 2016;280(3):793-804.

184. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in Radiologists' Interpretations of Mammograms. New England Journal of Medicine. 1994;331(22):1493-9.

185. Kanda T, Fukusato T, Matsuda M, Toyoda K, Oba H, Kotoku JI, et al. Gadolinium-based Contrast Agent Accumulates in the Brain Even in Subjects without Severe Renal Dysfunction: Evaluation of Autopsy Brain Specimens with Inductively Coupled Plasma Mass Spectroscopy. Radiology. 2015;276(1):228-32.

186. Turner EL, Metcalfe C, Donovan JL, Noble S, Sterne JAC, Lane JA, et al. Design and preliminary recruitment results of the Cluster randomised triAl of PSA testing for Prostate cancer (CAP). British Journal of Cancer. 2014;110(12):2829-36.

187. Centre TNI. Breast Screening Programme, England: 2015-16. The NHS Information Centre: Leeds. 2017.

188. Improvement N. National tariff payment system 2019/2020. Annex A: The national prices and national tariff workbook Accessed via https://improvementnhsuk/resources/national-tariff/. 2019.

189. Improvement N. National Schedule of Reference Costs 2017/18. National schedule of reference costs Accessed via https://improvementnhsuk/resources/reference-costs/. 2017.

190. Griebsch I, Brown J, Boggis C, Dixon A, Dixon M, Easton D, et al. Cost-effectiveness of screening with contrast enhanced magnetic resonance imaging vs X-ray mammography of women at a high familial risk of breast cancer. British Journal of Cancer. 2006;95(7):801-10.

191. Committee UNS. UK NSC recommendation on bowel cancer screening. UK NSC. 2018.

192. della Salute M. Screening per il cancro del colon-retto. Available from URL: http://wwwsalutegovit/portale/salute/p1_5jsp?lingua=italiano&id=24&area=Screening. 2018.

193. European guidelines for quality assurance in colorectal cancer screening and diagnosis: Overview and introduction to the full Supplement publication. Endoscopy. 2012;45(01):51-9.

194. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal Cancer Screening: Recommendations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorectal Cancer. The American journal of gastroenterology. 2017;112(7):1016-30.

195. Zdrowia M. Narodowy program zwalczania chorób nowotworowych. Założenia i cele operacyjne. 2006;2015.

196. Sonnenberg A. Cost-Effectiveness of Colonoscopy in Screening for Colorectal Cancer. Annals of Internal Medicine. 2000;133(8):573.

197. Quentin M, Blondin D, Arsov C, Schimmöller L, Hiester A, Godehardt E, et al. Prospective Evaluation of Magnetic Resonance Imaging Guided In-bore Prostate Biopsy versus Systematic Transrectal Ultrasound Guided Prostate Biopsy in Biopsy Naïve Men with Elevated Prostate Specific Antigen. The Journal of Urology. 2014;192(5):1374-9.

198. Rais-Bahrami S, Siddiqui MM, Vourganti S, Turkbey B, Rastinehad AR, Stamatakis L, et al. Diagnostic value of biparametric magnetic resonance imaging (MRI) as an adjunct to prostate-specific antigen (PSA)-based detection of prostate cancer in men without prior biopsies. BJU international. 2015;115(3):381-8.

199. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. New England Journal of Medicine. 1978;299(17):926-30.

200. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Radiology. 2003;226(1):24-8.

201. Barentsz JO, Weinreb JC, Verma S, Thoeny HC, Tempany CM, Shtern F, et al. Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use. European urology. 2016;69(1):41.

202. Padhani AR, Schoots I, Villeirs G. Contrast Medium or No Contrast Medium for Prostate Cancer Diagnosis. That Is the Question. Journal of Magnetic Resonance Imaging. 2020.

203. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, et al. ESUR prostate MR guidelines 2012. European radiology. 2012;22(4):746-57.

204. van der Leest M, Israël B, Cornel EB, Zámecnik P, Schoots IG, van der Lelij H, et al. High Diagnostic Performance of Short Magnetic Resonance Imaging Protocols for Prostate Cancer Detection in Biopsy-naïve Men: The Next Step in Magnetic Resonance Imaging Accessibility. European urology. 2019.

205. Bosaily AE-S, Frangou E, Ahmed HU, Emberton M, Punwani S, Kaplan R, et al. Additional Value of Dynamic Contrast-enhanced Sequences in Multiparametric Prostate Magnetic Resonance Imaging: Data from the PROMIS Study. European Urology. 2020.

206. Simmons LA, Ahmed HU, Moore CM, Punwani S, Freeman A, Hu Y, et al. The PICTURE study -- prostate imaging (multi-parametric MRI and Prostate HistoScanning) compared to transperineal ultrasound guided biopsy for significant prostate cancer risk evaluation. Contemporary clinical trials. 2014;37(1):69-83.

207. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ : British Medical Journal. 2015;351.

208. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS prostate imaging–reporting and data system: 2015, version 2. European urology. 2016;69(1):16-40.

209. Brizmohun Appayya M, Adshead J, Ahmed HU, Allen C, Bainbridge A, Barrett T, et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection–recommendations from a UK consensus meeting. BJU international. 2018;122(1):13-25.

210. Rosenkrantz AB, Lim RP, Haghighi 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. AJR Am J Roentgenol. 2013;201(4):W612-8.

211. Simmons LA, Kanthabalan A, Arya M, Briggs T, Barratt D, Charman SC, et al. Accuracy of transperineal targeted prostate biopsies, visual estimation and image fusion in men needing repeat biopsy in the picture trial. The Journal of urology. 2018;200(6):1227-34.

212. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.

213. R Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2013.

214. Padhani AR, Weinreb J, Rosenkrantz AB, Villeirs G, Turkbey B, Barentsz J. Prostate Imaging-Reporting and Data System Steering Committee: PI-RADS v2 Status Update and Future Directions. European Urology. 2019;75(3):385-96.

215. Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, Margolis D, et al. Reply to Erik Rud and Eduard Baco's Letter to the Editor re: Re: Jeffrey C. Weinreb, Jelle O. Barentsz, Peter L. Choyke, et al. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. Eur Urol 2016;69:16–40. European Urology. 2016;70(5):e137-e8.

216. Junker D, Steinkohl F, Fritz V, Bektic J, Tokas T, Aigner F, et al. Comparison of multiparametric and biparametric MRI of the prostate: are gadolinium-based contrast agents needed for routine examinations? World journal of urology. 2019;37(4):691-9.

217. Tan CH, Hobbs BP, Wei W, Kundra V. Dynamic contrast-enhanced MRI for the detection of prostate cancer: meta-analysis. AJR American journal of roentgenology. 2015;204(4):W439-48.

218. Haghighi M, Shah S, Taneja SS, Rosenkrantz AB. Prostate Cancer: Diffusion-Weighted Imaging Versus Dynamic-Contrast Enhanced Imaging for Tumor Localization—A Meta-Analysis. Journal of Computer Assisted Tomography. 2013;37(6):980-8.

219. Alabousi M, Salameh JP, Gusenbauer K, Samoilov L, Jafri A, Yu H, et al. Biparametric vs multiparametric prostate magnetic resonance imaging for the detection of prostate cancer in treatment-naïve patients: a diagnostic test accuracy systematic review and meta-analysis. Bju International. 2019;124(2):209-20.

220. Bass EJ, Pantovic A, Connor M, Gabe R, Padhani AR, Rockall A, et al. A systematic review and meta-analysis of the diagnostic accuracy of biparametric prostate MRI for prostate cancer in men at risk. Prostate Cancer and Prostatic Diseases. 2020.

221. Mertan FV, Greer MD, Shih JH, George AK, Kongnyuy M, Muthigi A, et al. Prospective Evaluation of the Prostate Imaging Reporting and Data System Version 2 for Prostate Cancer Detection. The Journal of Urology. 2016;196(3):690-6.

222. Taghipour M, Ziaei A, Alessandrino F, Hassanzadeh E, Harisinghani M, Vangel M, et al. Investigating the role of DCE-MRI, over T2 and DWI, in accurate PI-RADS v2 assessment of clinically significant peripheral zone prostate lesions as defined at radical prostatectomy. Abdominal Radiology. 2019;44(4):1520-7.

223. Sun C, Chatterjee A, Yousuf A, Antic T, Eggener S, Karczmar GS, et al. Comparison of T2-weighted imaging, DWI, and dynamic contrast-enhanced MRI for calculation of prostate cancer index lesion volume: correlation with whole-mount pathology. American Journal of Roentgenology. 2019;212(2):351-6.

224. Ben-Shlomo Y, Evans S, Ibrahim F, Patel B, Anson K, Chinegwundoh F, et al. The Risk of Prostate Cancer amongst Black Men in the United Kingdom: The PROCESS Cohort Study. European Urology. 2008;53(1):99-105.

225. Chu KC, Tarone RE, Freeman HP. Trends in prostate cancer mortality among black men and white men in the United States. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2003;97(6):1507-16.

226. Phillips AA, Jacobson JS, Magai C, Consedine N, Horowicz-Mehler NC, Neugut AI. Cancer incidence and mortality in the Caribbean. Cancer Investigation. 2007;25(6):476-83.

227. Bouchardy C, Verkooijen HM, Fioretta G. Social class is an important and independent prognostic factor of breast cancer mortality. International journal of cancer. 2006;119(5):1145-51.

228. Downs LS, Smith JS, Scarinci I, Flowers L, Parham G. The disparity of cervical cancer in diverse populations. Gynecologic oncology. 2008;109(2):S22-S30.

229. Carsin A, Sharp L, Cronin-Fenton D, Céilleachair AÓ, Comber H. Inequity in colorectal cancer treatment and outcomes: a population-based study. British journal of cancer. 2008;99(2):266-74.

230. Berglund A, Garmo H, Robinson D, Tishelman C, Holmberg L, Bratt O, et al. Differences according to socioeconomic status in the management and mortality in men with high risk prostate cancer. European Journal of Cancer. 2012;48(1):75-84.

231. Rapiti E, Fioretta G, Schaffar R, Neyroud-Caspar I, Verkooijen HM, Schmidlin F, et al. Impact of socioeconomic status on prostate cancer diagnosis, treatment, and prognosis. Cancer. 2009;115(23):5556-65.

232. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2015;65(2):87-108.

233. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. The Lancet. 2014;384(9959):2027-35.

234. Pinsky P, Miller A, Kramer B, Church T, Reding D, Prorok P, et al. Evidence of a Healthy Volunteer Effect in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. American Journal of Epidemiology. 2007;165(8):874-81.

235. Walsh EI, Turner EL, Lane JA, Donovan JL, Neal DE, Hamdy FC, et al. Characteristics of men responding to an invitation to undergo testing for prostate cancer as part of a randomised trial. Trials. 2016;17(1):497.

236. Kilpeläinen TP, Talala K, Raitanen J, Taari K, Kujala P, Tammela TLJ, et al. Prostate Cancer and Socioeconomic Status in the Finnish Randomized Study of Screening for Prostate Cancer. American Journal of Epidemiology. 2016;184(10):720-31.

237. Bossuyt P, Reitsma J, Bruns D, Gatsonis C, Glasziou P, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351: h5527. 2015.

238. Field JK, Duffy SW, Baldwin DR, Whynes DK, Devaraj A, Brain KE, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. Thorax. 2016;71(2):161-70.

239. Kummer S, Waller J, Ruparel M, Duffy SW, Janes SM, Quaife SL. Psychological outcomes of low-dose CT lung cancer screening in a multisite demonstration screening pilot: the Lung Screen Uptake Trial (LSUT). Thorax. 2020;75(12):1065-73.

240. Quaife SL, Ruparel M, Dickson JL, Beeken RJ, McEwen A, Baldwin DR, et al. Lung Screen Uptake Trial (LSUT): Randomized Controlled Clinical Trial Testing Targeted Invitation Materials. American journal of respiratory and critical care medicine. 2020;201(8):965-75.

241. Corbie-Smith G, Thomas SB, Williams MV, Moody-Ayers S. Attitudes and beliefs of African Americans toward participation in medical research. Journal of general internal medicine. 1999;14(9):537-46.

242. Weller DP, Patnick J, McIntosh HM, Dietrich AJ. Uptake in cancer screening programmes. The lancet oncology. 2009;10(7):693-9.

243. Duda C, Mahon I, Chen MH, Snyder B, Barr R, Chiles C, et al. Impact and costs of targeted recruitment of minorities to the National Lung Screening Trial. Clinical Trials. 2011;8(2):214-23.

244. Walsh EI, Turner EL, Lane JA, Donovan JL, Neal DE, Hamdy FC, et al. Characteristics of men responding to an invitation to undergo testing for prostate cancer as part of a randomised trial. Trials. 2016;17(1).

245. Thompson IM, Ankerst DP, Chi C, Goodman PJ, Tangen CM, Lucia MS, et al. Assessing Prostate Cancer Risk: Results from the Prostate Cancer Prevention Trial. JNCI: Journal of the National Cancer Institute. 2006;98(8):529-34.

246. Eldred-Evans D, Burak P, Connor MJ, Day E, Evans M, Fiorentino F, et al. Population-Based Prostate Cancer Screening With Magnetic Resonance or Ultrasonoagraphy: The IP1-PROSTAGRAM study. JAMA Oncology. 2021.

247. Hinshaw LB, Jackson SA, Chen MY. Direct mailing was a successful recruitment strategy for a lung-cancer screening trial. Journal of clinical epidemiology. 2007;60(8):853-7.

248. Roobol MJ, Kranse R, Bangma CH, Van Leenders AGJLH, Blijenberg BG, Van Schaik RHN, et al. Screening for Prostate Cancer: Results of the Rotterdam Section of the European Randomized Study of Screening for Prostate Cancer. European Urology. 2013;64(4):530-9.

249. Grönberg H, Adolfsson J, Aly M, Nordström T, Wiklund P, Brandberg Y, et al. Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. The lancet oncology. 2015;16(16):1667-76.

250. Field JK, Duffy SW, Baldwin DR, Brain KE, Devaraj A, Eisen T, et al. The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. Health Technology Assessment. 2016;20(40):1-146.

251. Virtanen A, Anttila A, Luostarinen T, Malila N, Nieminen P. Improving cervical cancer screening attendance in Finland. International Journal of Cancer. 2015;136(6):E677-E84.

252. Weinrich SP, Reynolds Jr WA, Tingen MS, Starr CR. Barriers to prostate cancer screening. Cancer Nursing. 2000;23(2):117-21.

253. Nijs H, Essink-Bot M-L, DeKoning H, Kirkels W, Schröder F. Why do men refuse or attend population-based screening for prostate cancer? Journal of public health. 2000;22(3):312-6.

254. Toledano MB, Smith RB, Brook JP, Douglass M, Elliott P. How to establish and follow up a large prospective cohort study in the 21st century-lessons from UK COSMOS. PLoS One. 2015;10(7):e0131521.

255. Brizmohun Appayya M, Adshead J, Ahmed HU, Allen C, Bainbridge A, Barrett T, et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection - recommendations from a UK consensus meeting. BJU Int. 2018;122(1):13-25.

256. Naing L, Winn T, Rusli B. Practical issues in calculating the sample size for prevalence studies. Archives of orofacial Sciences. 2006;1:9-14.

257. Schoots IG. MRI in early prostate cancer detection: how to manage indeterminate or equivocal PI-RADS 3 lesions? Translational Andrology and Urology. 2018;7(1):70-82.

258. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤4.0 ng per Milliliter. New England Journal of Medicine. 2004;350(22):2239-46.

259. de Rooij M, Hamoen EHJ, Fütterer JJ, Barentsz JO, Rovers MM. Accuracy of Multiparametric MRI for Prostate Cancer Detection: A Meta-Analysis. American Journal of Roentgenology. 2014;202(2):343-51.

260. Lane JA, Donovan JL, Davis M, Walsh E, Dedman D, Down L, et al. Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial. The Lancet Oncology. 2014;15(10):1109-18.

261. Gwet K. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. Statistical methods for inter-rater reliability assessment. 2002;1(6):1-6.

262. Altman DG. Practical statistics for medical research: CRC press; 1990.

263. Nam RK, Wallis CJ, Stojcic-Bendavid J, Milot L, Sherman C, Sugar L, et al. A Pilot Study to Evaluate the Role of Magnetic Resonance Imaging for Prostate Cancer Screening in the General Population. J Urol. 2016;196(2):361-6.

264. Stamey TA, Freiha FS, McNeal JE, Redwine EA, Whittemore AS, Schmid HP. Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. Cancer. 1993;71(3 Suppl):933-8.

265. Miah S, Eldred-Evans D, Simmons LA, Shah TT, Kanthabalan A, Arya M, et al. Patient reported outcome measures for transperineal template prostate mapping biopsies in the PICTURE study. The Journal of urology. 2018;200(6):1235-40.

266. Morgan TO, Jacobsen SJ, McCarthy WF, Jacobson DJ, McLeod DG, Moul JW. Age-Specific Reference Ranges for Serum Prostate-Specific Antigen in Black Men. New England Journal of Medicine. 1996;335(5):304-10.

267. Kulasingam SL, Hughes JP, Kiviat NB, Mao C, Weiss NS, Kuypers JM, et al. Evaluation of Human Papillomavirus Testing in Primary Screening for Cervical Abnormalities. Jama. 2002;288(14):1749.

268. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. Jama. 2012;307(13):1394-404.

269. De Groot JAH, Janssen KJM, Zwinderman AH, Bossuyt PMM, Reitsma JB, Moons KGM. Correcting for Partial Verification Bias: A Comparison of Methods. Annals of Epidemiology. 2011;21(2):139-48.

270. Catalona WJ, Hudson MLA, Scardino PT, Richie JP, Ahmann FR, Flanigan RC, et al. Selection of optimal prostate specific antigen cutoffs for early detection of prostate cancer: receiver operating characteristic curves. The Journal of urology. 1994;152(6 Part 1):2037-42.

271. Alonzo TA, Pepe MS. Using a combination of reference tests to assess the accuracy of a new diagnostic test. Statistics in medicine. 1999;18(22):2987-3003.

272. Harel O, Zhou XH. Multiple imputation for correcting verification bias. Statistics in medicine. 2006;25(22):3769-86.

273. Diamond GA, Rozanski A, Forrester JS, Morris D, Pollock BH, Staniloff HM, et al. A model for assessing the sensitivity and specificity of tests subject to selection bias. Journal of Chronic Diseases. 1986;39(5):343-55.

274. Merl T, Scholz M, Gerhardt P, Langer M, Laubenberger J, Weiss HD, et al. Results of a prospective multicenter study for evaluation of the diagnostic quality of an open whole-body low-field MRI unit. European Journal of Radiology. 1999;30(1):43-53.

275. Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM. Effect of Verification Bias on Screening for Prostate Cancer by Measurement of Prostate-Specific Antigen. New England Journal of Medicine. 2003;349(4):335-42.

276. Lauer MS. [18F]Fluorodeoxyglucose Uptake by Positron Emission Tomography for Diagnosis of Suspected Lung Cancer. Archives of Internal Medicine. 2007;167(2):161.

277. Van Der Heijden GJMG, T. Donders AR, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. Journal of Clinical Epidemiology. 2006;59(10):1102-9.

278. Kang Z, Min X, Weinreb J, Li Q, Feng Z, Wang L. Abbreviated Biparametric Versus Standard Multiparametric MRI for Diagnosis of Prostate Cancer: A Systematic Review and Meta-Analysis. American Journal of Roentgenology. 2018;212(2):357-65.

279. Flicker L, Logiudice D, Carlin JB, Ames D. The predictive value of dementia screening instruments in clinical populations. International Journal of Geriatric Psychiatry. 1997;12(2):203-9.

280. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz J. Spectrum bias in the evaluation of diagnostic tests: Lessons from the rapid dipstick test for urinary tract infection. Annals of Internal Medicine. 1992;117(2):135-40.

281. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Statistics in medicine. 1997;16(9):981-91.

282. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. bmj. 2016;353.

283. Martin RM, Donovan JL, Turner EL, Metcalfe C, Young GJ, Walsh EI, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. Jama. 2018;319(9):883-95.

284. Humphrey LL, Helfand M, Chan BS, Woolf SH. Breast cancer screening: A summary of the evidence for the u.s. preventive services task force. Annals of Internal Medicine. 2002;137(5_Part_1):347-60.

285. Kriege M, Brekelmans CTM, Boetes C, Besnard PE, Zonderland HM, Obdeijn IM, et al. Efficacy of MRI and Mammography for Breast-Cancer Screening in Women with a Familial or Genetic Predisposition. New England Journal of Medicine. 2004;351(5):427-37.

286. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography. Annals of internal medicine. 2011;155(8):481-92.

287. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. Acta Radiologica. 2000;41(1):52-6.

288. Atkins BL, Athanasou N, Deeks JJ, Crook DW, Simpson H, Peto TE, et al. Prospective evaluation of criteria for microbiological diagnosis of prosthetic-joint infection at revision arthroplasty. The OSIRIS Collaborative Study Group. J Clin Microbiol. 1998;36(10):2932-9.

289. Vernon SW. Participation in Colorectal Cancer Screening: A Review. JNCI Journal of the National Cancer Institute. 1997;89(19):1406-22.

290. Heijnsdijk EAM, Adolfsson J, Auvinen A, Roobol MJ, Hugosson J, De Koning HJ. The Impact of Design and Performance in Prostate-Specific Antigen Screening: Differences Between ERSPC Centers. European Urology. 2019;76(3):276-9.

291. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. American Journal of Gastroenterology. 2008;103(6):1541-9.

292. Vart G, Banzi R, Minozzi S. Comparing participation rates between immunochemical and guaiac faecal occult blood tests: A systematic review and meta-analysis. Preventive Medicine. 2012;55(2):87-92.

293. Hol L, de Jonge V, Van Leerdam M, van Ballegooijen M, Looman C, Van Vuuren A, et al. Screening for colorectal cancer: comparison of perceived test burden of guaiac-based faecal occult blood test, faecal immunochemical test and flexible sigmoidoscopy. European journal of cancer. 2010;46(11):2059-66.

294. Svensson MH, Svensson E, Lasson A, Hellström M. Patient acceptance of CT colonography and conventional colonoscopy: prospective comparative study in patients with or suspected of having colorectal disease. Radiology. 2002;222(2):337-45.

295. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical care. 2003:582-92.

296. Powell AA, Burgess DJ, Vernon SW, Griffin JM, Grill JP, Noorbaloochi S, et al. Colorectal cancer screening mode preferences among US veterans. Preventive Medicine. 2009;49(5):442-8.

297. Carlsson S, Aus G, Wessman C, Hugosson J. Anxiety associated with prostate cancer screening with special reference to men with a positive screening test (elevated PSA) – Results from a prospective, population-based, randomised study. European Journal of Cancer. 2007;43(14):2109-16.

298. Brindle LA, Oliver SE, Dedman D, Donovan JL, Neal DE, Hamdy FC, et al. Measuring the psychosocial impact of population-based prostate-specific antigen testing for prostate cancer in the UK. BJU international. 2006;98(4):777-82.

299. de Wijkerslooth TR, de Haan MC, Stoop EM, Bossuyt PM, Thomeer M, Essink-Bot M-L, et al. Burden of colonoscopy compared to non-cathartic CT-colonography in a colorectal cancer screening programme: randomised controlled trial. Gut. 2012;61(11):1552-9.

300. Raz DJ, Nelson RA, Kim JY, Sun V. Pilot study of a video intervention to reduce anxiety and promote preparedness for lung cancer screening. Cancer Treatment and Research Communications. 2018;16:1-8.

301. Gareen IF, Duan F, Greco EM, Snyder BS, Boiselle PM, Park ER, et al. Impact of lung cancer screening results on participant health-related quality of life and state anxiety in the National Lung Screening Trial. Cancer. 2014;120(21):3401-9.

302. Pooler BD, Baumel MJ, Cash BD, Moawad FJ, Riddle MS, Patrick AM, et al. Screening CT colonography: multicenter survey of patient experience, preference, and potential impact on adherence. American journal of roentgenology. 2012;198(6):1361-6.

303. Udayakumar N, Porter KK. How Fast Can We Go: Abbreviated Prostate MR Protocols. Current Urology Reports. 2020;21(12).

304. 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.

305. Ling BS, Moskowitz MA, Wachs D, Pearson B, Schroy III PC. Attitudes toward colorectal cancer screening tests: a survey of patients and physicians. Journal of general internal medicine. 2001;16(12):822-30.

306. van Gelder RE, Birnie E, Florie J, Schutter MP, Bartelsman JF, Snel P, et al. CT Colonography and Colonoscopy: Assessment of Patient Preference in a 5-week Follow-up Study. Radiology. 2004;233(2):328-37.

307. Deutekom M, Terra MP, Dijkgraaf MGW, Dobben AC, Stoker J, Boeckxstaens GE, et al. Patients' perception of tests in the assessment of faecal incontinence. The British Journal of Radiology. 2006;79(938):94-100.

308. Feinstein AR. XXXI. On the sensitivity, specificity, and discrimination of diagnostic tests. Clinical Pharmacology & Therapeutics. 1975;17(1):104-16.

309. Gosselaar C, Roobol MJ, Roemeling S, Van Der Kwast TH, Schröder FH. Screening for prostate cancer at low PSA range: The impact of digital rectal examination on tumor incidence and tumor characteristics. The Prostate. 2007;67(2):154-61.

310. Vickers AJ, Cronin AM, Bjork T, Manjer J, Nilsson PM, Dahlin A, et al. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. BMJ. 2010;341(sep14 1):c4521-c.

311. Grenabo Bergdahl A, Wilderäng U, Aus G, Carlsson S, Damber J-E, Frånlund M, et al. Role of Magnetic Resonance Imaging in Prostate Cancer Screening: A Pilot Study Within the Göteborg Randomised Screening Trial. European Urology. 2016;70(4):566-73.

312. von Karsa L, Patnick J, Segnan N, Atkin W, Halloran S, Lansdorp-Vogelaar I, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. Endoscopy. 2013;45(1):51.

313. Khoo CC, Eldred-Evans D, Peters M, Bertoncelli Tanaka M, Noureldin M, Miah S, et al. Likert vs PI-RADS v2: a comparison of two radiological scoring systems for detection of clinically significant prostate cancer. BJU International. 2019.

314. Sankoh AJ, D'Agostino Sr RB, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. Statistics in Medicine. 2003;22(20):3133-50.

315. Pagniez MA, Kasivisvanathan V, Puech P, Drumez E, Villers A, Olivier J. Predictive Factors of Missed Clinically Significant Prostate Cancers in Men with Negative Magnetic Resonance Imaging: A Systematic Review and Meta-Analysis. Journal of Urology. 2020;204(1):24-32.

316. Ahdoot M, Wilbur AR, Reese SE, Lebastchi AH, Mehralivand S, Gomella PT, et al. MRI-Targeted, Systematic, and Combined Biopsy for Prostate Cancer Diagnosis. New England Journal of Medicine. 2020;382(10):917-28.

317. NICE. NICE Guidance - Prostate cancer: diagnosis and management. BJU International. 2019;124(1):9-26.

318. England N. The NHS long term plan. 2019.

319. Montgomery M, McCrone SH. Psychological distress associated with the diagnostic phase for suspected breast cancer: systematic review. Journal of Advanced Nursing. 2010;66(11):2372-90.

320. Lal A, Phillips S, Russell C, Woolhouse I. The novel use of fast track CT to select patients for lung cancer clinics: effect on clinic efficiency, waiting times, and patient satisfaction. 2011;87(1026):264-8.

321. Beggs AD, Bhate RD, Irukulla S, Achiek M, Abulafi AM. Straight to colonoscopy: the ideal patient pathway for the 2-week suspected cancer referrals? The Annals of The Royal College of Surgeons of England. 2011;93(2):114-9.

322. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. bmj. 2015;350:h2750.

323. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. International journal of technology assessment in health care. 2003;19(4):613-23.

324. Tamada T, Sone T, Jo Y, Yamamoto A, Yamashita T, Egashira N, et al. Prostate cancer: relationships between postbiopsy hemorrhage and tumor detectability at MR diagnosis. Radiology. 2008;248(2):531-9.

325. Qayyum A, Coakley FV, Lu Y, Olpin JD, Wu L, Yeh BM, et al. Organ-confined prostate cancer: effect of prior transrectal biopsy on endorectal MRI and MR spectroscopic imaging. American Journal of Roentgenology. 2004;183(4):1079-83.

326. Kirkham A, Haslam P, Keanie J, McCafferty I, Padhani A, Punwani S, et al. Prostate MRI: who, when, and how? Report from a UK consensus meeting. Clinical radiology. 2013;68(10):1016-23.

327. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Academic pediatrics. 2013;13(6):S38-S44.

328. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research: Ravenio Books; 2015.

329. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics. 2002;27(4):299-309.

330. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. Journal of clinical pharmacy and therapeutics. 2002;27(4):299-309.

331. UK National Health Service (NHS) Health Research Authority. Is my study research? [Available from: http://www.hra-decisiontools.org.uk/ethics/index.html.

332. Barrett T, Slough R, Sushentsev N, Shaida N, Koo BC, Caglic I, et al. Three-year experience of a dedicated prostate mpMRI pre-biopsy programme and effect on timed cancer diagnostic pathways. Clinical Radiology. 2019;74(11):894.e1-.e9.

333. Delaloge S, Bonastre J, Borget I, Garbay J-R, Fontenay R, Boinon D, et al. The challenge of rapid diagnosis in oncology: Diagnostic accuracy and cost analysis of a large-scale one-stop breast clinic. European Journal of Cancer. 2016;66:131-7.

334. Kum F, Elhage O, Maliyil J, Wong K, Faure Walker N, Kulkarni M, et al. Initial outcomes of local anaesthetic freehand transperineal prostate biopsies in the outpatient setting. BJU international. 2020;125(2):244-52.

335. Wetterauer C, Shahin O, Federer-Gsponer JR, Keller N, Wyler S, Seifert HH, et al. Feasibility of freehand MRI/US cognitive fusion transperineal biopsy of the prostate in local anaesthesia as in-office procedure—experience with 400 patients. Prostate Cancer and Prostatic Diseases. 2020;23(3):429-34.

336. Drost F-JH, Osses D, Nieboer D, Bangma CH, Steyerberg EW, Roobol MJ, et al. Prostate Magnetic Resonance Imaging, with or Without Magnetic Resonance Imaging-targeted Biopsy, and Systematic Biopsy for Detecting Prostate Cancer: A Cochrane Systematic Review and Meta-analysis. European Urology. 2020;77(1):78-94.

337. Maringe C, Spicer J, Morris M, Purushotham A, Nolte E, Sullivan R, et al. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. The Lancet Oncology. 2020;21(8):1023-34.

338. Faria R, Soares MO, Spackman E, Ahmed HU, Brown LC, Kaplan R, et al. Optimising the diagnosis of prostate cancer in the era of multiparametric magnetic resonance imaging: a cost-effectiveness analysis based on the Prostate MR Imaging Study (PROMIS). European urology. 2018;73(1):23-30.

339. Drudge-Coates L, Khati V, Ballesteros R, Martyn-Hemphill C, Brown C, Green J, et al. A nurse practitioner model for the assessment of suspected prostate cancer referrals is safe, cost and time efficient. Ecancermedicalscience. 2019;13:994-.

340. Schoots IG, Padhani AR. Personalizing prostate cancer diagnosis with multivariate risk prediction tools: how should prostate MRI be incorporated? World journal of urology. 2020;38(3):531-45.

341. Schröder F, Kattan MW. The Comparability of Models for Predicting the Risk of a Positive Prostate Biopsy with Prostate-Specific Antigen Alone: A Systematic Review. 2008;54(2):274-90.

342. Radtke JP, Wiesenfarth M, Kesch C, Freitag MT, Alt CD, Celik K, et al. Combined Clinical Parameters and Multiparametric Magnetic Resonance Imaging for Advanced Risk Modeling of Prostate Cancer—Patient-tailored Risk Stratification Can Reduce Unnecessary Biopsies. European Urology. 2017;72(6):888-96.

343. Mehralivand S, Shih JH, Rais-Bahrami S, Oto A, Bednarova S, Nix JW, et al. A Magnetic Resonance Imaging–Based Prediction Model for Prostate Biopsy Risk Stratification. JAMA Oncology. 2018;4(5):678.

344. Perez IM, Jambor I, Kauko T, Verho J, Ettala O, Falagario U, et al. Qualitative and Quantitative Reporting of a Unique Biparametric MRI: Towards Biparametric MRI-Based Nomograms for Prediction of Prostate Biopsy Outcome in Men With a Clinical Suspicion of Prostate Cancer (IMPROD and MULTI-IMPROD Trials). Journal of Magnetic Resonance Imaging. 2020;51(5):1556-67.

345. Alberts AR, Roobol MJ, Verbeek JFM, Schoots IG, Chiu PK, Osses DF, et al. Prediction of High-grade Prostate Cancer Following Multiparametric Magnetic Resonance Imaging: Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculators. European Urology. 2019;75(2):310-8.

346. Distler FA, Radtke JP, Bonekamp D, Kesch C, Schlemmer H-P, Wieczorek K, et al. The value of PSA density in combination with PI-RADS™ for the accuracy of prostate cancer prediction. The Journal of urology. 2017;198(3):575-82.

347. Connor M, Eldred-Evans D, van Son M, Hosking-Jervis F, Bertoncelli Tanaka M, Reddy D, et al. A Multicentre Study of the Clinical Utility of Non-targeted Systematic Transperineal

Prostate Biopsies in Patients Undergoing Pre-biopsy mpMRI. The Journal of Urology. 2020:10.1097/JU. 0000000000001184.

348. Merisaari H, Jambor I, Ettala O, Boström PJ, Montoya Perez I, Verho J, et al. IMPROD biparametric MRI in men with a clinical suspicion of prostate cancer (IMPROD Trial): Sensitivity for prostate cancer detection in correlation with whole-mount prostatectomy sections and implications for focal therapy. Journal of Magnetic Resonance Imaging. 2019;50(5):1641-50.

349. Jambor I, Boström PJ, Taimen P, Syvänen K, Kähkönen E, Kallajoki M, et al. Novel biparametric MRI and targeted biopsy improves risk stratification in men with a clinical suspicion of prostate cancer (IMPROD Trial). 2017.

350. Dell'Oglio P, Stabile A, Soligo M, Brembilla G, Esposito A, Gandaglia G, et al. There Is No Way to Avoid Systematic Prostate Biopsies in Addition to Multiparametric Magnetic Resonance Imaging Targeted Biopsies. European Urology Oncology. 2020;3(1):112-8.

351. Stabile A, Dell'Oglio P, De Cobelli F, Esposito A, Gandaglia G, Fossati N, et al. Association Between Prostate Imaging Reporting and Data System (PI-RADS) Score for the Index Lesion and Multifocal, Clinically Significant Prostate Cancer. European Urology Oncology. 2018.

352. Miah S, Hosking-Jervis F, Connor MJ, Eldred-Evans D, Shah TT, Arya M, et al. A Multicentre Analysis of the Detection of Clinically Significant Prostate Cancer Following Transperineal Image-fusion Targeted and Nontargeted Systematic Prostate Biopsy in Men at Risk. European Urology Oncology. 2020;3(3):262-9.

353. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj. 2009;338.

354. Janssen KJ, Donders ART, Harrell Jr FE, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. Journal of clinical epidemiology. 2010;63(7):721-7.

355. Buuren Sv, Groothuis-Oudshoorn K, Robitzsch A, Vink G, Doove L, Jolani S. Multivariate imputation by chained equations–Package ‚mice '. Computer software] http://cran r-project org/web/packages/mice. 2012.

356. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons; 2004.

357. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Statistical methods in medical research. 2017;26(2):796-808.

358. Fowler JE, Bigler SA, Kilambi NK, Land SA. Relationships between prostate-specific antigen and prostate volume in black and white men with benign prostate biopsies. 1999;53(6):1175-8.

359. Park JJ, Chen M-H, Loffredo M, D'Amico AV. Prostate-Specific Antigen Velocity Before and After Elimination of Factors That Can Confound the Prostate-Specific Antigen Level. 2012;82(3):1217-21.

360. Steyerberg EW. Clinical prediction models: Springer; 2019.

361. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837-45.

362. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;15(4):361-87.

363. Sullivan LM, Massaro JM, D'Agostino Sr RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Statistics in medicine. 2004;23(10):1631-60.

364. Schoots IG, Padhani AR, Rouvière O, Barentsz JO, Richenberg J. Analysis of Magnetic Resonance Imaging–directed Biopsy Strategies for Changing the Paradigm of Prostate Cancer Diagnosis. European urology oncology. 2020;3(1):32-41.

365. Kranse R, Roobol M, Schröder FH. A graphical device to represent the outcomes of a logistic regression analysis. The prostate. 2008;68(15):1674-80.

366. Thompson IM, Leach RJ, Ankerst DP. Focusing PSA testing on detection of high-risk prostate cancers by incorporating patient preferences into decision making. Jama. 2014;312(10):995-6.

367. Rentsch C, Bebu I, Guest JL, Rimland D, Agan BK, Marconi V. Combining epidemiologic and biostatistical tools to enhance variable selection in HIV cohort analyses. PloS one. 2014;9(1):e87352.

368. McCray GPJ, Titman AC, Ghaneh P, Lancaster GA. Sample size re-estimation in paired comparative diagnostic accuracy studies with a binary response. BMC Medical Research Methodology. 2017;17(1).

369. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. Journal of clinical epidemiology. 1988;41(10):949-58.

370. Delgado R, Tibau X-A. Why Cohen's Kappa should be avoided as performance measure in classification. PloS one. 2019;14(9):e0222916.

371. Girometti R, Giannarini G, Greco F, Isola M, Cereser L, Como G, et al. Interreader agreement of PI-RADS v. 2 in assessing prostate cancer with multiparametric MRI: A study using whole-mount histology as the standard of reference. Journal of Magnetic Resonance Imaging. 2019;49(2):546-55.

372. Care DoHaS. NHS Payment by Results 2010-11 National Tariff Information. Accessed via https://datagovuk. 2014.

373. Pastorino U, Bellomi M, Landoni C, De Fiori E, Arnaldi P, Picchio M, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. The Lancet. 2003;362(9384):593-7.

374. Graser A, Stieber P, Nagel D, Schafer C, Horst D, Becker CR, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. Gut. 2009;58(2):241-8.

375. Whyte S, Thomas C, Kearns B, Webster M, Chilcott J. Optimising Bowel Cancer Screening Phase 1: Optimising the cost effectiveness of repeated FIT screening and screening strategies combining bowel scope and FIT screening. 2017.

376. Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. Results of initial low-dose computed tomographic screening for lung cancer. The New England journal of medicine. 2013;368(21):1980-91.

377. Porté F, Uppara M, Malietzis G, Faiz O, Halligan S, Athanasiou T, et al. CT colonography for surveillance of patients with colorectal cancer: Systematic review and meta-analysis of diagnostic efficacy. 2016.

378. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, Lamonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA: A Cancer Journal for Clinicians. 2018;68(4):250-81.

379. Sharaf RN, Ladabaum U. Comparative effectiveness and cost-effectiveness of screening colonoscopy vs. sigmoidoscopy and alternative strategies. The American journal of gastroenterology. 2013;108(1):120-32.

380. ONS. Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2018: Office for National Statistics; 2018 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2018.

381. Richards M. The independent review of adult screening programmes in England. England: NHS. 2019.

382. Heijnsdijk E, De Carvalho T, Auvinen A, Zappa M, Nelen V, Kwiatkowski M, et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. Journal of the National Cancer Institute. 2015;107(1):dju366.

383. Woo S, Suh CH, Kim SY, Cho JY, Kim SH, Moon MH. Head-to-Head Comparison Between Biparametric and Multiparametric MRI for the Diagnosis of Prostate Cancer: A Systematic Review and Meta-Analysis. American Journal of Roentgenology. 2018;211(5):W226-W41.

384. Woo S, Suh CH, Kim SY, Cho JY, Kim SH, Moon MH. Interactive Up-to-date Meta analysis on Biparametric vs. Multiparametric MRI for the Diagnosis of Prostate Cancer [Available from: http://www.iu-ma.org/prostate/bpmpmri/.

385. Munn Z, Moola S, Lisy K, Riitano D, Murphy F. Claustrophobia in magnetic resonance imaging: A systematic review and meta-analysis. Radiography. 2015;21(2):e59-e63.

386. Iwan E, Yang J, Enders J, Napp AE, Rief M, Dewey M. Patient preferences for development in MRI scanner design: a survey of claustrophobic patients in a randomized study. European Radiology. 2020.

387. Nordström T, Jäderling F, Carlsson S, Aly M, Grönberg H, Eklund M. Does a novel diagnostic pathway including blood-based risk prediction and MRI-targeted biopsies outperform prostate cancer screening using prostate-specific antigen and systematic prostate biopsies? - protocol of the randomised study STHLM3MRI. BMJ Open. 2019;9(6):e027816.

388. Carlsson S, Assel M, Sjoberg D, Ulmert D, Hugosson J, Lilja H, et al. Influence of blood prostate specific antigen levels at age 60 on benefits and harms of prostate cancer screening: population based cohort study. BMJ. 2014;348(mar28 1):g2296-g.

389. Andriole GL, Levin DL, Crawford ED, Gelmann EP, Pinsky PF, Chia D, et al. Prostate Cancer Screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: findings from the initial screening round of a randomized trial. Journal of the National Cancer Institute. 2005;97(6):433-8.

390. Postma R, Schröder FH, Van Leenders GJLH, Hoedemaeker RF, Vis AN, Roobol MJ, et al. Cancer Detection and Cancer Characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC) – Section Rotterdam. European Urology. 2007;52(1):89-97.

391. D'Orsi C, Bassett L, Feig S. Breast imaging reporting and data system (BI-RADS). Breast Imaging Lee CI, Lehman CD, Bassett LW (ed): Oxford University Press, New York. 2018.

392. Kweldam CF, Kümmerlin IP, Nieboer D, Verhoef EI, Steyerberg EW, Incrocci L, et al. Prostate cancer outcomes of men with biopsy Gleason score 6 and 7 without cribriform or intraductal carcinoma. 2016;66:26-33.

393. Kweldam CF, Kümmerlin IP, Nieboer D, Steyerberg EW, Bangma CH, Incrocci L, et al. Presence of invasive cribriform or intraductal growth at biopsy outperforms percentage grade 4 in predicting outcome of Gleason score 3+4=7 prostate cancer. Modern Pathology. 2017.

394. Panebianco V, Barchetti G, Simone G, Del Monte M, Ciardi A, Grompone MD, et al. Negative multiparametric magnetic resonance imaging for prostate cancer: what's next? European urology. 2018;74(1):48-54.

395. Houlahan KE, Salmasi A, Sadun TY, Pooli A, Felker ER, Livingstone J, et al. Molecular Hallmarks of Multiparametric Magnetic Resonance Imaging Visibility in Prostate Cancer. European Urology. 2019;76(1):18-23.

396. Akin O, Riedl CC, Ishill NM, Moskowitz CS, Zhang J, Hricak H. Interactive dedicated training curriculum improves accuracy in the interpretation of MR imaging of prostate cancer. European radiology. 2010;20(4):995-1002.

397. Rawashdeh MA, Lee WB, Bourne RM, Ryan EA, Pietrzyk MW, Reed WM, et al. Markers of good performance in mammography depend on number of annual readings. Radiology. 2013;269(1):61-7.

398. Scott HJ, Gale AG. Breast screening: PERFORMS identifies key mammographic training needs. The British journal of radiology. 2006;79(special_issue_2):S127-S33.

399. Balleyguier C, Ayadi S, Van Nguyen K, Vanel D, Dromain C, Sigal R. BIRADS™ classification in mammography. European Journal of Radiology. 2007;61(2):192-4.

400. Gage JC, Schiffman M, Katki HA, Castle PE, Fetterman B, Wentzensen N, et al. Reassurance Against Future Risk of Precancer and Cancer Conferred by a Negative Human Papillomavirus Test. JNCI Journal of the National Cancer Institute. 2014;106(8):dju153-dju.

401. Wentzensen N, Clarke MA, Bremer R, Poitras N, Tokugawa D, Goldhoff PE, et al. Clinical Evaluation of Human Papillomavirus Screening With p16/Ki-67 Dual Stain Triage in a Large Organized Cervical Cancer Screening Program. JAMA Internal Medicine. 2019;179(7):881.

402. Ronco G, Dillner J, Elfström KM, Tunesi S, Snijders PJ, Arbyn M, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. The lancet. 2014;383(9916):524-32.

403. Catalona WJ, D'Amico AV, Fitzgibbons WF, Kosoko-Lasaki O, Leslie SW, Lynch HT, et al. What the U.S. Preventive Services Task Force Missed in Its Prostate Cancer Screening Recommendation. Annals of Internal Medicine. 2012;157(2):137.

404. Walter FM, Thompson MJ, Wellwood I, Abel GA, Hamilton W, Johnson M, et al. Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. BMC Cancer. 2019;19(1).

405. ReIMAGINE Prostate Cancer Screening. https://ClinicalTrials.gov/show/NCT04063566.

406. Prostate Cancer Screening Trial Using Imaging. https://ClinicalTrials.gov/show/NCT03702439.

407. Independent UKPoBCS. The benefits and harms of breast cancer screening: an independent review. The Lancet.380(9855):1778-86.

408. Ronco G, Dillner J, Elfström KM, Tunesi S, Snijders PJF, Arbyn M, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. The Lancet.383(9916):524-32.

409. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. The American journal of gastroenterology. 2008;103(6):1541-9.

Appendix I: Research Approvals for IP1-PROSTAGRAM

Integrated Research Application System
Application Form for Other clinical trial or investigation

The Chief Investigator should complete this form. Guidance on the questions is available wherever you see this symbol displayed. We recommend reading the guidance first. The complete guidance and a glossary are available by selecting Help.

Please define any terms or acronyms that might not be familar to lay reviewers of the application.

Short title and version number: (maximum 70 characters - this will be inserted as header on all forms)
PROSTAGRAM – Image-based prostate cancer testing in the community

PART A: Core study information

1. ADMINISTRATIVE DETAILS

A1. Full title of the research:

PROSTAGRAM - Prostate Screening Trial using A Group of Radiological Approaches including MRI and ultrasound

A2-1. Educational projects

Name and contact details of student(s):

Student 1

Title	Forename/Initials	Surname
Mr	David	Eldred-Evans

Address	5L15 Laboratory Block Charing Cross Campus Imperial College London
Post Code	W6 8RP
E-mail	d.eldred-evans@imperial.ac.uk
Telephone	0203 311 5471
Fax	

Give details of the educational course or degree for which this research is being undertaken:

Name and level of course/ degree:
PhD

Name of educational establishment:
Imperial College London

Name and contact details of academic supervisor(s):

Academic supervisor 1

Title	Forename/Initials	Surname
Professor	Hashim	Ahmed

NHS
Health Research Authority

London - Camberwell St Giles Research Ethics Committee
Level 3, Block B
Whitefriars
Lewins Mead
Bristol
BS1 2NT

Tel: 0207 104 8059

Please note: This is the favourable opinion of the REC only and does not allow the amendment to be implemented at NHS sites in England until the outcome of the HRA assessment has been confirmed.

06 June 2019

Mr David Eldred-Evans
5L15 Laboratory Block
Charing Cross Campus
Imperial College London
W6 8RP

Dear Mr Eldred-Evans

Study title:	**PROSTAGRAM - Prostate Screening Trial using A Group of Radiological Approaches including MRI and ultrasound**
REC reference:	**18/LO/1338**
Protocol number:	**18HH4595**
Amendment number:	**2**
Amendment date:	**01 May 2019**
IRAS project ID:	**247728**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Notice of Substantial Amendment (non-CTIMP) [Amendment_Form_SA2_V5.11_01MAY2019_signed.pdf]	2	01 May 2019
Research protocol or project proposal [Protocol_Version_V1.2_01MAY2019.pdf]	1.2	01 May 2019

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

Working with NHS Care Organisations

Sponsors should ensure that they notify the R&D office for the relevant NHS care organisation of this amendment in line with the terms detailed in the categorisation email issued by the lead nation for the study.

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.

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: https://www.hra.nhs.uk/planning-and-improving-research/learning/

18/LO/1338:	**Please quote this number on all correspondence**

Yours sincerely

PP
Mr John Richardson
Chair

Appendix II: Worksheets for IP1-PROSTAGRAM

MRI Report Worksheet

PROSTAGRAM

Subject ID Number:

P 1 - ☐☐☐

Size of Prostate

Transverse	AP	Cranio-Caudal
☐.☐cm	☐.☐cm	☐.☐cm

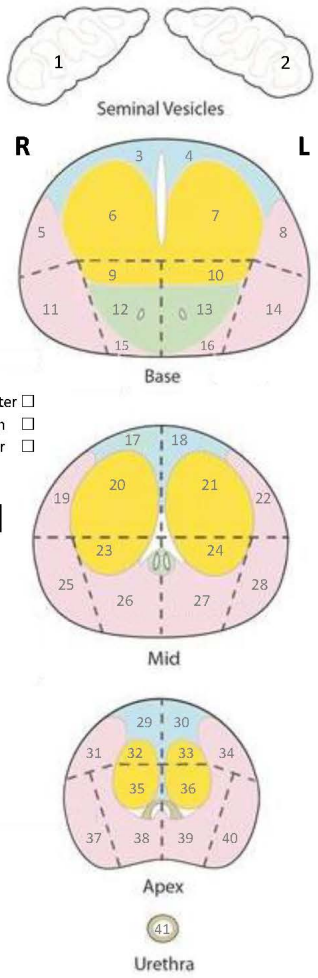
Scan Date d d m m y y

Individual Lesions

Are there any lesions with an MRI score ≥ 3? Yes ☐ No ☐

(Please draw & number lesions on diagram and complete table)

	WITHOUT CAD					WITH CAD	
Lesion	**Rank the Severity (1-4)**	**Max Dia-meter (mm)**	**Modified PIRADS Score**	**LIKERT Score**	**Focal or Diffuse (if Score = 3)**	**CAD Positive?**	**CAD Score**
MRI Lesion					Focal ☐ Diffuse ☐	Yes ☐ No ☐	
MRI Lesion					Focal ☐ Diffuse ☐	Yes ☐ No ☐	
MRI Lesion					Focal ☐ Diffuse ☐	Yes ☐ No ☐	
MRI Lesion					Focal ☐ Diffuse ☐	Yes ☐ No ☐	



Radiological stage if subsequently found to have cancer

≤T2 ☐ T2a ☐ T2b ☐ T2c ☐ T3a ☐ T3b ☐ T4 ☐ Sphincter ☐ Rectum ☐ Bladder ☐

Here radiological T3a/b means unequivocal extracapsular disease.

Any incidental findings? None ☐ Bladder Tumour ☐ Rectal Tumour ☐

Other: _____

Is there a reason why the MRI needs to be repeated?

None	☐	Geometric distortion	☐	Susceptibility	☐
Internal motion	☐	Signal graininess	☐	Blurring	☐
External motion	☐	Ghosting	☐	Other: _____	

Any additional/new lesions on CAD? Yes ☐ No ☐ Excluded ☐

Lesion	Maximum Diameter (mm)	Modified PIRADS Score	LIKERT
New CAD Target			
New CAD Target			

I declare that I have completed the initial review blinded to the CAD output.

Signature / Initials	Reporter Name

Version 1.3 22 NOV 2018 1

US Report Worksheet

PROSTAGRAM

Subject ID Number:

P 1 - ☐☐☐

Is there an US to report? Yes ☐ No ☐

Scan Date d d m m y y

Size of Prostate

Transverse ☐.☐cm AP ☐.☐cm Cranio-Caudal ☐.☐cm

Is there significant calcification present causing an acoustic shadow? Yes ☐ No ☐

IF YES: Location of the calcification ☐

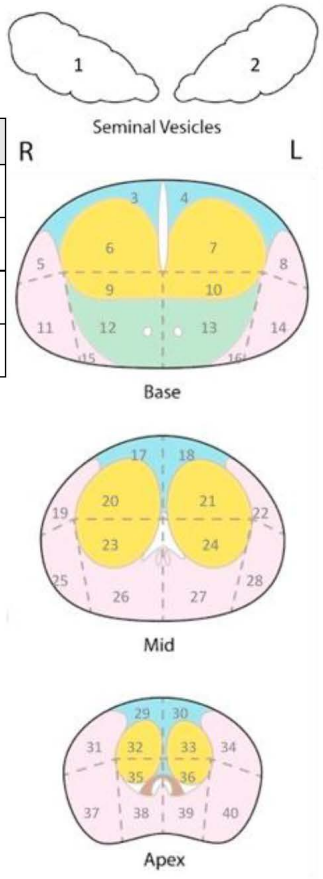
Is the anterior portion of the prostate visible? Yes ☐ No ☐

Individual Lesions

Are there any lesions with an US score ≥ 3? Yes ☐ No ☐

(Please draw & number lesions on diagram and complete table)

Lesion	E_{Ratio}	E_{mean}	E_{Max}	SD	Diam	Dept	US Score	If score = 3 Doppler	Dopp score
US Lesion 1								+ve ☐ -ve ☐	
US Lesion 2								+ ve ☐ - ve ☐	
US Lesion 3								+ ve ☐ - ve ☐	
US Lesion 4								+ ve ☐ - ve ☐	



Radiological stage if subsequently found to have cancer

≤T2 ☐ T2a ☐ T2b ☐ T2c ☐ T3a ☐ T3b ☐ T4 ☐

Here radiological T3a/b means unequivocal extracapsular disease.

Any incidental findings? None ☐ Bladder Tumour ☐ Rectal Tumour ☐

Other: _____

Signature

Reporter Name:

Appendix III: Pre-Screening and Information Sheet

Participant Information Sheet

Version: 1.3; 17 December 2018
Sponsor protocol number: 18HH4595
REC reference: 18/LO/1338
Sponsor: Imperial College London

PROSTAGRAM
Prostate Screening Trial using A Group of Radiological Approaches including MRI and ultrasound

PATIENT INFORMATION SHEET

We would like to invite you to take part in a research study testing a new prostate health check. Please read this booklet and think about whether you would like to take part. Talk about it with family and friends. Ask us if there is anything that is not clear or you would like more information about. We will go through this patient information sheet with you and answer any questions you might have. You do not have to decide straight away.

PART 1 tells you the purpose of this study and what will happen if you choose to take part

PART 2 gives you more detailed information about the conduct of the study

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not to take part.

PART 1

1. WHAT IS THE PURPOSE OF THE PROSTAGRAM STUDY?

The PROSTAGRAM study will test a new prostate health check-up in men aged 50 to 69. Prostate problems are common in men, particularly in those over the age of about 50 years. We aim to find a combination of tests which may spot prostate diseases or conditions early, often before you notice anything, when treatment could be simpler and more successful.

The main prostate condition that we are looking for is prostate cancer. Prostate cancer often grows slowly and has a low risk of spreading but some prostate cancers grow more quickly and can shorten a man's life. When it is found early the majority of cases can be cured. Not all prostate cancers need treatment. Some can be safely monitored to make sure they are not changing. This is because many men above the age of 50 will have tiny low risk prostate cancers which will never grow or spread. However, some prostate cancers can be aggressive and need to be found early so they can be treated. When treatment is needed there are lots of options to choose from if the cancer is contained in the prostate.

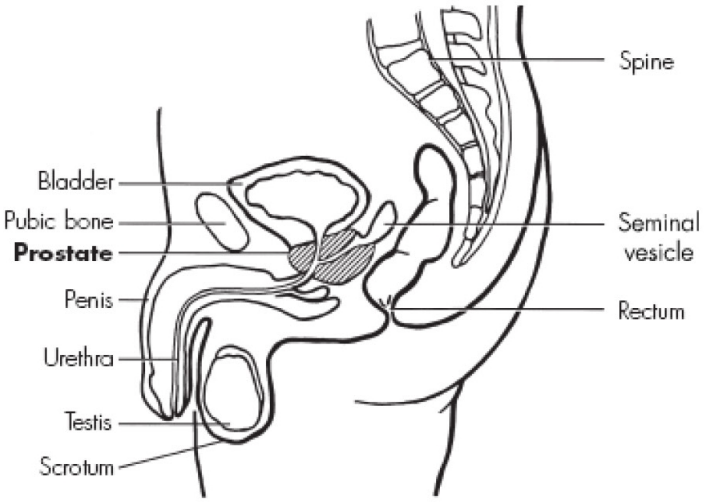
Prostate cancer is currently diagnosed using a blood test called prostate specific antigen (PSA) done by the GP. If the level of PSA is high then men are referred to hospital to have biopsies. PSA testing can find cancers early but because the PSA is not specific to cancer it cannot tell the difference between cancers that need treatment and those that can be monitored. In other cancers, such as breast cancer, there are imaging tests like mammograms which are offered to everyone at a certain age with the aim of finding tumours early. There are no equivalent types of tests for men and prostate cancer.

There are specialised scans which are used in hospitals to help find prostate cancer more accurately in men who have already been referred by their GP. The PROSTAGRAM study will be testing if it is possible to use these scans in the community to find problems at an early stage when treatment can be more successful. Men who agree to take part will be invited to a prostate health check to have a PSA blood test and these specialised prostate scans.

2. WHAT IS THE PROSTATE?

The prostate is an organ that forms part of the male reproductive system. It is located immediately below the bladder, just in front of the bowel. Its main function is to produce fluid that makes up part of the semen. In younger men the prostate is about the size of a walnut. It surrounds the beginning of the urethra, the tube that takes urine from the bladder and through the penis.

Figure 1: Location of the prostate



3. Why have I been asked to join the PROSTAGRAM study?

If you have received a letter from your GP or been approached verbally it is because your GP is one of several practices who have agreed to invite suitable patients to this study. From medical records, we think that you may be suitable to receive the tests for the prostate health check.

Alternatively, you may have heard about the PROSTAGRAM study in another way. We are hoping to recruit about 300 to 400 men to take part so we are contacting many people like you in the local area.

4. Am I Eligible for the study?

There are a number of things we look at to see if you are eligible for this study. You may be able to take part in the PROSTAGRAM study if you:

- Are aged between 50 and 69 years
- Able to have an MRI and rectal ultrasound of the prostate

A small number of men will not be able to take part in the study. This includes those who

- Have had a PSA test in last 2 years
- Would be too ill from other conditions or diseases to have treatment for prostate cancer even if we did find it
- Are not able to decide for themselves if they want to take part in the study

We will go through these in greater detail in the first trial visit if you wish to participate.

4. Do I have to take part?

No, it's up to you. We will go through the information about the study with you in the health check clinic. If you decide to take part we will ask you to sign a consent form saying that you are happy to take part in the study.

If you do decide to part you can change your mind later and you do not have to tell us why. It will not make a difference to the usual care you get from your doctor.

5. What will happen to me if I decide to take part?

You will attend a prostate health check clinic at Imperial College or Imperial College NHS Trust. When you arrive for your visit a member of the research team will see you. They will explain the study to you and discuss any questions or concerns that you may have. Once you have talked about the study and if you decide to take part you will be asked to sign the study consent form.

A member of the study team will ask you to complete some questionnaires about your health. You will be asked to give a sample of blood to measure your PSA level. If you have signs of a urine infection, you may be asked for a sample of urine to look for infection. A doctor will perform an examination and ultrasound scan of your prostate. You will also have an MRI scan of your prostate, which will be done at one of our imaging centres.

The screening visit and imaging studies will be held on a single day or split across more than one-day dependent on your preference and availability of clinic and MRI time. You will be contacted to receive your results within 6 weeks of having the tests. If all the tests are negative, there will be no further study visits.

If any of these tests are positive, you will be offered a prostate biopsy and an appointment to see a specialist in the hospital which will be done in person.

6. What are the different tests in PROSTAGRAM?

There are three tests in PROSTAGRAM. These include PSA blood test and two different types of scans. These have previously only been available for men suspected of having prostate cancer who have been referred to hospital.

PSA

PSA is a protein produced by both normal cells and cancer cells in the prostate. Men usually have a small amount of PSA in their blood, and the amount rises as you get older and your prostate gets bigger. Inflammation or infection of the prostate can lead to high PSA tests.

Although a raised PSA level can be a sign of prostate cancer, many men with a raised PSA level don't actually have prostate cancer. Alternatively, some men with a normal PSA level can also have prostate cancer.

Prostate ultrasound

Prostate ultrasound uses sound waves to build up a picture of the prostate. These sound waves are safe and are the same as those used in a pregnant woman to see the baby in the womb. The ultrasound machine shows us what the prostate looks like and also how hard different areas of the prostate are. Prostate cancer tends to be harder or denser than normal tissue. The test involves passing a small ultrasound device into the back passage (rectum). It

is done at the same time as an examination of the prostate using a gloved finger. You may find this test uncomfortable but it should not be painful.

MRI

An MRI (magnetic resonance imaging) scan creates a detailed picture of the prostate using magnetic waves. It is the most common scan used at the moment to look for prostate cancer. An MRI is a safe procedure. It does not use x-rays or radiation. Before the scan you will be given an injection of a medication to relax the bowel. This helps to reduce the movement of your bowels and makes the pictures clearer.

During the scan, you will be asked to lie on a padded table which gently moves you into the MRI scanner. You will need to lie still for about 20 minutes. The machine will make loud, thumping and whirring noises, much like the sound of a washing machine. Although it is a painless test, some people may find it a little noisy. Some men who have claustrophobia (fear of enclosed spaces) may not be able to tolerate the scan. There is a 2-way microphone within the scanner so that you will be able to talk to the staff. The staff will also be able to see you on a monitor and provide reassurance if there are any concerns.

7. WHAT WILL HAPPEN AFTER GETTING THE TESTS

You will be contacted with your results within 6 weeks of having the tests.

Negative Tests

If the tests are negative it means that it is highly unlikely that you have a fast-growing prostate cancer. It is important to remember that:

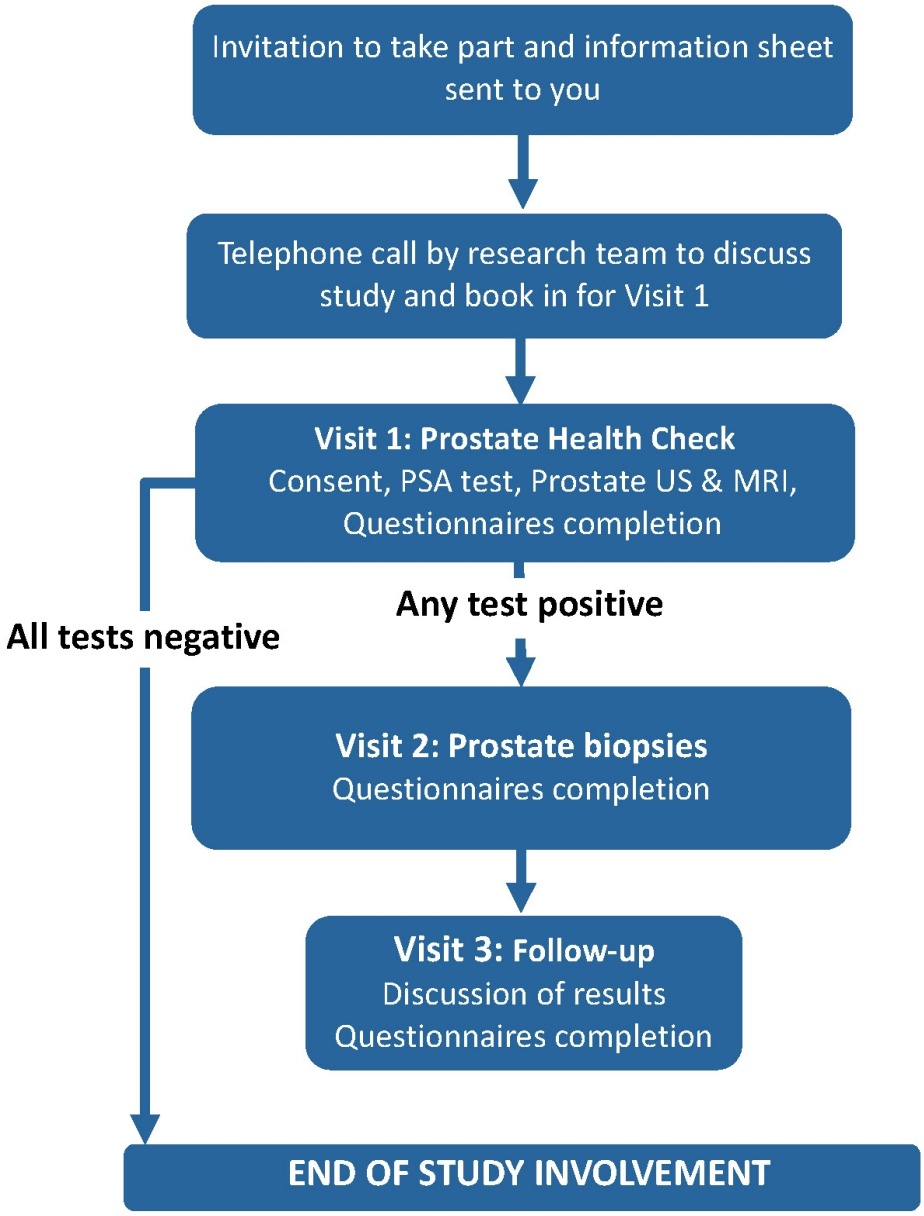
- A negative test does not mean you will never get prostate cancer. It is important that you see your GP if you have any concerns about this in future
- It is important to remember that these tests are for prostate cancer, so cannot rule out cancers elsewhere in the body

Positive Tests

If any of your tests are positive, it does not mean you definitely have prostate cancer. If a test is positive, you will:

- Be offered a prostate biopsy which involves using thin needles to take small pieces of tissue from the prostate. The tissue is then looked at under a microscope to check for cancer
- We will not be able to tell you which of the three tests is positive until the end of the trial
- The result of the prostate biopsy will be sent to your GP. You will be offered an appointment with the specialist urology team at Imperial College Healthcare NHS Trust to discuss the results

FIGURE 1: DIAGRAM OF PROSTAGRAM



8. What happens if I need a prostate biopsy?

You will receive local anaesthetic to numb the area of the biopsies. An ultrasound scanning probe will be passed into the back passage to allow the clinician to see your prostate during the biopsy. Prostate biopsies are taken by passing a needle through the skin between your back passage and the base of your penis. This is called the transperineal route. Any areas of your prostate that looked suspicious on either the ultrasound or the MRI will be highlighted on a screen. The surgeon can see these and take samples of them with the biopsy needle as well as sampling other parts of the prostate. If there are no areas of suspicion on the scans, but the PSA test is high, a routine set of prostate biopsies will be taken. This is because the imaging tests may have missed something. The clinician taking the biopsies will take them in a prearranged order from any abnormal areas detected inside the prostate. The order of the biopsies (ultrasound-targeted or MRI-targeted first) will be assigned at random by a computer. This is an attempt to make the results of the study more accurate.

Your prostate biopsy samples will be stored in Imperial College Healthcare NHS Trust pathology department for a period of 30 years and then destroyed as per the standard operating regulations of the NHS laboratories.

9. What else could I be asked to do?

There are some additional optional research requests.

- If you decide to sign the optional portion of the consent form, we will ask you to provide extra blood and urine samples to be collected and stored for research (up to 50 ml of blood and up to 250 ml of urine).

  If you take part in PROSTAGRAM, we would like your permission to use these stored samples for prostate cancer research. These research studies are not expected to benefit you, but may help to improve the diagnosis and/or the treatment of prostate cancer for future patients.

  Any extra blood and urine samples that you give us for these research studies will be stored securely in Imperial College Healthcare Biobanks for a period of 10 years so that we can repeat any tests on them if necessary, and use them to look at new tests for prostate cancer. These samples will be identified using a special study number assigned to you, in such a way that the scientists analysing them will not be able to find out your identity. This research would be carried out only after approval from an independent research ethics committee and would involve extracting DNA or other chemicals from the samples to see whether the tests make it is easier to detect or monitor the effect of treatment for prostate cancer. These samples would be considered a gift from you and no personal results from these tests or studies could be provided to you.

- We will also ask if you are happy to have an additional blood sample for the Episwitch biomarker panel. This is an optional test which investigates how certain genes related to prostate cancer are switched on or switched off by factors outside of the genetic code, a process known as epigenetics. This test could help us develop an epigenetic-based prognostic test for prostate cancer. The anonymised samples will be sent to Oxford Biodynamics (OBD) Reference Laboratory in the UK.

- Imaging scans are performed as part of this study. We would also like to know if you are willing for us to store your imaging scans and use your scan data to see if new ways of looking at these scans can detect cancer better in the future.

- We will also ask you if you are happy to give consent for your health status to be followed up over time. This will be done by linking your name and NHS number with records held by the NHS and maintained by the NHS Information Centre and the NHS Central Register or any applicable NHS information system. This will allow us to track what happens after the study finishes to see if anyone gets cancer in the future and about the type of cancer and the treatment they have had. Results of your optional health status check will also help us to refer to any future upcoming studies. This does not mean that that you will need to make any follow up visits.

- We would also like to know if you are willing for us to record and store your partial postcode. This part of the study is also optional. Your partial postcode will be collected at study registration, then coded and kept confidential in a secure password protected database. The partial postcode will be used to study the performance of the recruitment for the study across different areas.

10. What are the potential benefits of taking part?

You may have prostate cancer picked up more quickly than it would have been if you were not taking part in the study. This could mean that the cancer is more treatable and the chance of surviving is better. However, there is no guarantee that taking part in this study will benefit you personally.

You will also provide important information to help us find out if these imaging tests might work in the community. We will use this information to decide whether a bigger study is needed and how best to design that study. By taking part you are playing an essential role in research and could be helping future generations.

11. What are the possible disadvantages and risks in taking part?

There are a few things you should be aware of in terms of the possible downsides of taking part in this study.

First, we might find slow-growing or non-aggressive cancers that might not cause any symptoms or problems in your lifetime. You would have a discussion with a doctor and then have to decide whether to have treatment or whether to have your cancer monitored. Treatment can cause side effects that can be hard to live with. But having your cancer monitored rather than having treatment might make you worry about your cancer.

Second, some men and their families may get very anxious and worried while they are waiting for their tests results.

Third, with any tests for cancer there can be false negatives and false positives. No medical test is completely accurate. We think these tests will pick up the majority of fast-growing prostate cancers. However, they may not pick up all cases of aggressive prostate cancer (false negative). The tests can sometimes be positive in men who do not have aggressive prostate cancer (false positive). These men will be offered a prostate biopsy which can sometimes cause side-effects, such as discomfort, urine infection, difficulty urinating, and blood in the urine, bowel movements or semen. These side-effects are temporary.

Finally, like all medicines, the medications used during the MRI scan can sometimes cause side effects. These include blurred vision, dry mouth, dizziness, increased heart rate, constipation and pain at the injection site. These side-effects usually wear off by the time the scan has finished, but if you experience blurred vision, we advise you not to drive or operate machinery until this has worn off. These medications may very rarely cause acute glaucoma (high pressure in the eyeball). If you develop a painful eye in the 24 hours after the injection you should attend A&E immediately.

12. WHAT HAPPENS WHEN THE STUDY STOPS?

Once you receive the results of your tests at the final study visit, you will either return to your GP's care or be looked after by the clinical team in hospital.

It is also important for us to know how you are doing even after your participation in the study has stopped so we can follow up on your health status to help future related research. For this reason, we will ask for your consent for your name to be used to gather information from records held by the NHS and maintained by the NHS Information Centre and the NHS Central Register or any applicable NHS information system (including linkage to routine hospital admission data). In order for us to do this we provide identifiable information for us to trace you on the National Health Service Care Register (NHSCR) for up to 10 years (this is an optional part of the study).

13. CAN I CHANGE MY MIND?

Yes, you can decide not to have any of the procedures at any time. Depending on when you change your mind, your doctor will recommend that you continue with standard care.

This completes Part 1 of the information sheet. If you are considering participating in the study, please continue to read the additional information in Part 2 before making your decision.

PART 2

1. WHAT HAPPENS IF RELEVANT NEW INFORMATION BECOMES AVAILABLE?

Data from this study will be monitored regularly by researchers who are independent of the study. Sometimes, during the course of a research project, new information becomes available about the procedures that are being studied. If you are in the study and this happens, your study doctor will tell you about it and discuss with you whether you want to, or should, continue in the study. If you decide not to carry on, your study doctor will make arrangements for your care to continue. If you decide to continue in the study you will be asked to sign a consent form that includes new information. Also, on receiving new information your study doctor might consider it to be in your best interests to stop the medical procedures in the study. If so they will explain the reasons and arrange for your medical treatment to continue another way. If the study is stopped for any other reason, you will be told why and your doctor will arrange for your continuing treatment.

2. WHAT IF SOMETHING GOES WRONG?

Every care will be taken in the course of this study. However in the unlikely event that you are injured by taking part, compensation may be available.

Imperial College London holds insurance policies which apply to this study. If you experience harm or injury as a result of taking part in this study, you will be eligible to claim compensation without having to prove that Imperial College is at fault. This does not affect your legal rights to seek compensation.

If you are harmed due to someone's negligence, then you may have grounds for a legal action. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been treated during the course of this study, then you should immediately inform the Chief Investigator (Professor Ahmed on 0203 311 5473 or via email on hashim.ahmed@nhs.net) The normal National Health Service mechanisms are also available to you including the Patient Advice and Liaison Service (PALS) who can be contacted on 020 3312 7777. If you are still not satisfied with the response, you may contact the Imperial College, Joint Research Compliance Office.

3. WHAT WILL HAPPEN IF I LOSE THE CAPACITY TO CONSENT DURING THE TRIAL?

In the unlikely event that you lose capacity to consent during your participation in the study, you would be withdrawn from the study. Any data or tissue already collected with consent would be retained and used in the study. No further data or tissue would be collected.

Imperial College London is the sponsor for this study based in the United Kingdom. We will be using information from you and your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. Imperial College London will keep identifiable information about you from this study 10 years after the study has finished in relation to data subject consent forms.

Further information on Imperial College London's retention periods may be found at https://www.imperial.ac.uk/media/imperial-college/administration-and-support-services/records-and-archives/public/RetentionSchedule.pdf.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.
You can find out more about how we use your information by contacting Professor Hashim Ahmed via e-mail at Hashim.ahmed@nhs.net.

LEGAL BASIS

As a university we use personally-identifiable information to conduct research to improve health, care and services. As a publicly-funded organisation, we have to ensure that it is in the public interest when we use personally-identifiable information from people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use your data in the ways needed to conduct and analyse the research study.
Health and care research should serve the public interest, which means that we have to demonstrate that our research serves the interests of society as a whole. We do this by following the UK Policy Framework for Health and Social Care Research.

INTERNATIONAL TRANSFERS

There may be a requirement to transfer information to countries outside the European Economic Area (for example, to a research partner). Where this information contains your personal data, Imperial College London will ensure that it is transferred in accordance with data protection legislation. If the data is transferred to a country which is not subject to a European Commission (EC) adequacy decision in respect of its data protection standards, Imperial College London will enter into a data sharing agreement with the recipient organisation that incorporates EC approved standard contractual clauses that safeguard how your personal data is processed.

CONTACT US

If you wish to raise a complaint on how we have handled your personal data or if you want to find out more about how we use your information, please contact Imperial College London's Data Protection Officer via email at dpo@imperial.ac.uk, via telephone on 020 7594 3502 and via post at Imperial College London, Data Protection Officer, Faculty Building Level 4, London SW7 2AZ.

If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO). The ICO does recommend that you seek to resolve matters with the data controller (us) first before involving the regulator.

5. WHO IS ORGANISING AND FUNDING THE STUDY?

PROSTAGRAM is funded by a Wellcome Trust Senior Clinical Research Fellowship awarded to Professor Ahmed and Imperial College London (grant code: 204998/Z/16/Z). Further funding is from the Urology Foundation awarded to Dr Eldred-Evans and Imperial College London.

The study is being carried out by Imperial College London. This research is being undertaken as part of a PhD degree at the Department of Surgery and Cancer, Imperial College London by Dr Eldred-Evans. The researchers include a team of specialised doctors, scientists, technical staff and nurses. Our team is experienced and has conducted similar research in the field.

6. WHO HAS REVIEWED THE STUDY?

Patients and expert reviewers have looked at the study both before the funding was awarded and after. This study was given a favourable ethical opinion for conduct in the NHS by the Camberwell St Giles Research Ethics Committee. This committee is responsible for making sure that research takes place in a way that protects the patients' rights and welfare.

7. Will I get paid for taking part?

There is no payment for taking part in this study. Reasonable travel expenses can be reimbursed and we will offer refreshments if your appointment is over lunch.

8. WHAT WILL HAPPEN TO THE RESULTS OF THE RESEARCH STUDY?

If you would like to know our overall findings of the study on all the men who took part, we can send you the final study results. Large studies such as this take a few years to complete and for the final results to appear. When the study results are concluded, we will publish our findings in scientific journals and present them at scientific meetings. It will not be possible to identify you in the published results.

Pre-Screening Telephone Script

Prostate Health Check

NHS
Imperial College Healthcare
NHS Trust

Study Information (GP referral)

'Hello, This is _____ calling from Hammersmith Hospital I am calling regarding the Prostate Health Check. [I apologise for the delay in contacting you we've been receiving a large number of calls about this prostate health check.]

Is this a good time to speak or would you like one of the team to call you back at a different time?

How do your hear about the health check? Did you get a letter from you GP / What was the name of your GP Surgery?

If not from GP letter then go back and select the correct option *

- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ [redacted]
- ○ Other (please specify):

The health check involves 3 simple tests: A blood test, An MRI and An examination of the prostate with an ultrasound. The reason we are doing this is that although the vast majority of men your age will have a healthy prostate, after age 50 is also when a few men start to develop prostate cancer and that's what these tests will look for.

OPTION 1: IF BOOKING FOR CLINICAL IMAGING FACILITY (HAMMERSMITH)
It is being run at Hammersmith hospital on Wednesdays and you will get all these tests over about 2 hours.
Is this something that would interest you?

OPTION 2: IF BOOKING FOR CHARING CROSS + MOUNT VERNON

Due to high demand we are offering appointments at Charing Cross Hospital on Fridays. You will get the blood test and ultrasound over about 2 hours. The MRI scan will happen at Mount Vernon Hospital and you will receive an appointment for this when we see you at Charing Cross Hospital.

Is this something that would interest you?

- [ ] Yes
- [ ] No

To give you a bit more information about test. The first test is a simple blood test, called PSA (prostate specific antigen), which looks at the health of the prostate. Have you had a blood test before? *(respond appropriately)*

The second test you will meet a doctor who will examine the prostate with an ultrasound. This is involving passing a small ultrasound device into the back passage. This device looks just the first couple of centimeters into the back passage as that is where the prostate is.

The last test is an MRI scan which takes 20 minutes. Before the MRI scan we give a injection of a safe medication which will allow us to see the prostate better. An MRI does not use any radiation and works using magnets. It is very rare for there to be any issues but there have been rare cases of overheating and burns to skin. This tends to happen if your skin is a bit sweaty and moist during the scan.

Did you have any questions about any of these tests? Do you understand that the ultrasound will be passed into the bottom *(be aware of uncertainty about ultrasound)*

Does it sound like you would comfortable with all these tests?

At the start and at the end of each test you be asked to complete a number of questionnaires about each test. Although we have been using these tests for many years as specialists to look for the prostate cancer this is the first time we are offering them to everyone in the local area. So this is part of research study and we hope that if it is successful we can roll these out to everyone across the country. So we need to get your feedback on how you found each test.

Would you like me to book an appointment? If so I just need to check a few details

- ◯ Yes
- ◯ Other (please specify):

Previous Page | Next Page

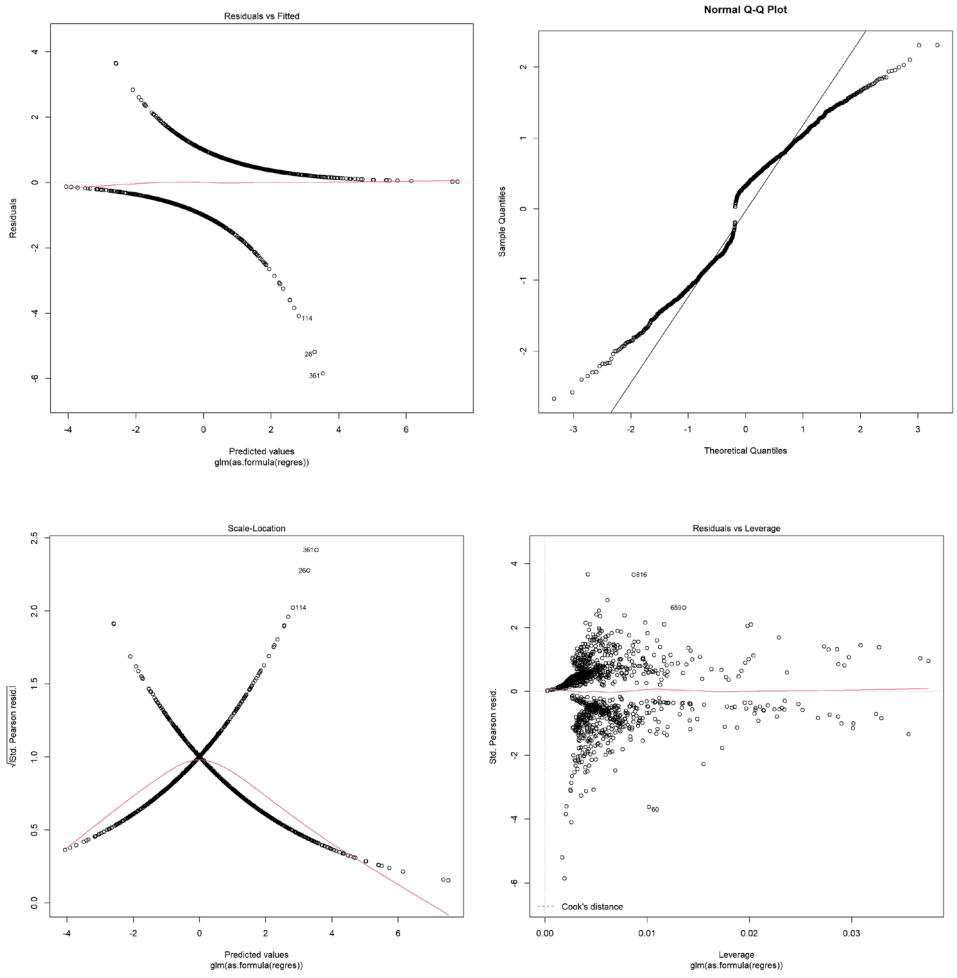
Appendix IV: RAPID Risk Model Details

Gleason ≥3+4

Model 1 (simple) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-1.857579	0.655685	-2.833	0.00461 **
Age	0.052708	0.009823	5.366	8.06e-08 ***
PSA Density (log)	1.106359	0.132386	8.357	< 2e-16 ***
Prior Biopsy	-1.023608	0.334249	-3.062	0.00220 **
MRI Volume	-0.008101	0.003220	-2.516	0.01187 *
MRI Score 4	0.935536	0.184440	5.072	3.93e-07 ***
MRI Score 5	1.891128	0.201489	9.386	< 2e-16 ***

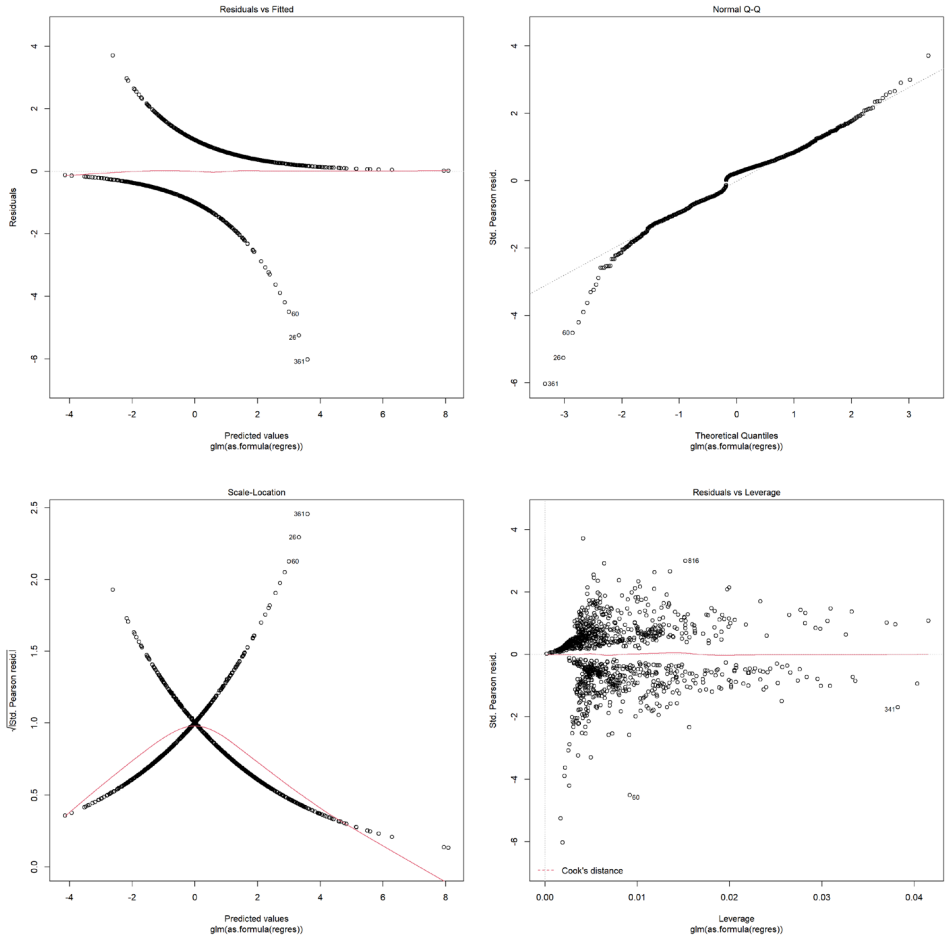
Model 1 (simple) Regression Diagnostics



Model 2 (non-invasive) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-2.696226	0.747821	-3.605	0.000312 ***
Age	0.065402	0.011113	5.885	3.97e-09 ***
PSA Density (log)	1.122160	0.133989	8.375	< 2e-16 ***
Prior Biopsy	-1.048128	0.338159	-3.100	0.001938 **
MRI Volume	-0.007841	0.003229	-2.428	0.015164 *
MRI Score 4	0.953666	0.186409	5.116	3.12e-07 ***
MRI Score 5	1.906551	0.203102	9.387	< 2e-16 ***
Afro-Caribbean	4.202472	1.592930	2.638	0.008335 **
Family History	0.385932	0.215076	1.794	0.072750
Interaction term	-0.071797	0.024740	-2.902	0.003707 **

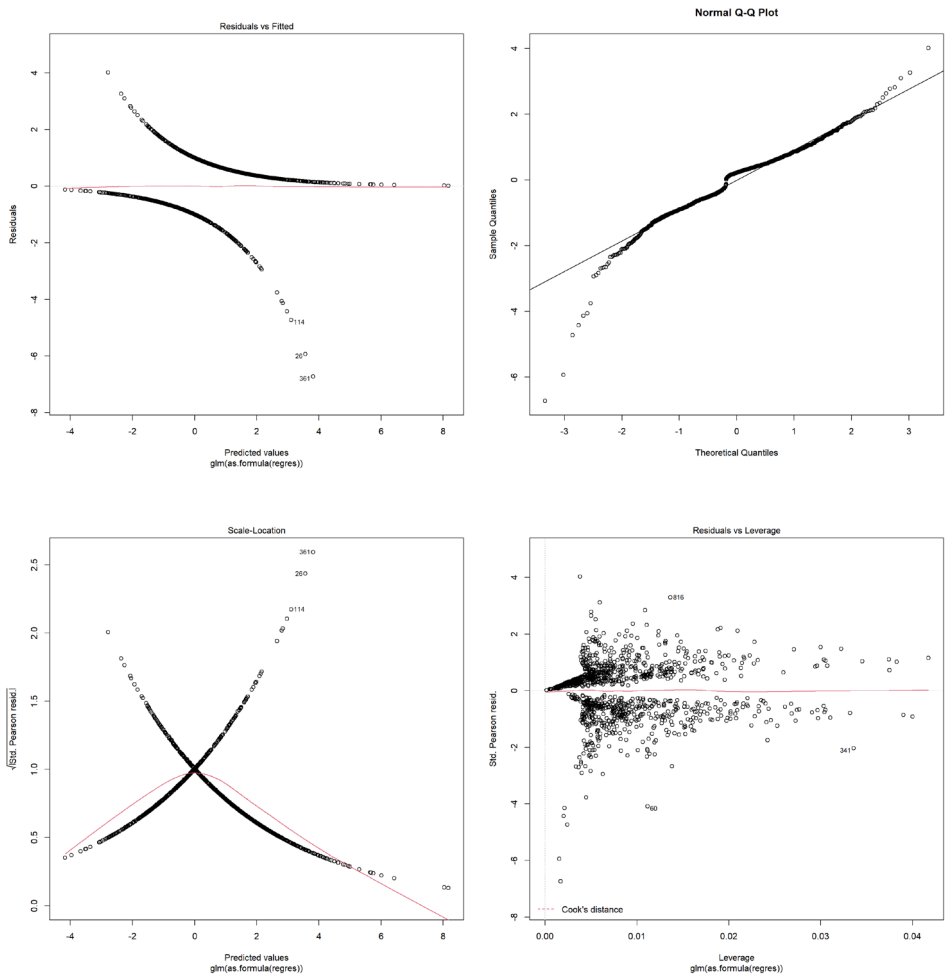
Model 2 (non-invasive) Regression Diagnostics



Model 3 (Full) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-2.789757	0.755342	-3.693	0.000221 ***
Age	0.063571	0.011238	5.657	1.54e-08 ***
PSA Density (log)	1.089768	0.134823	8.083	6.32e-16 ***
Prior Biopsy	-0.914319	0.338504	-2.701	0.006912 **
MRI Volume	-0.008173	0.003265	-2.503	0.012308 *
MRI Score 4	0.944250	0.187352	5.040	4.66e-07 ***
MRI Score 5	1.808195	0.204609	8.837	< 2e-16 ***
Afro-Caribbean	4.061677	1.600220	2.538	0.011142 * .
Family History	0.388261	0.216840	1.791	0.073367.
DRE Abnormal	0.561151	0.153492	3.656	0.000256 ***

Model 3 (Full) Regression Diagnostics

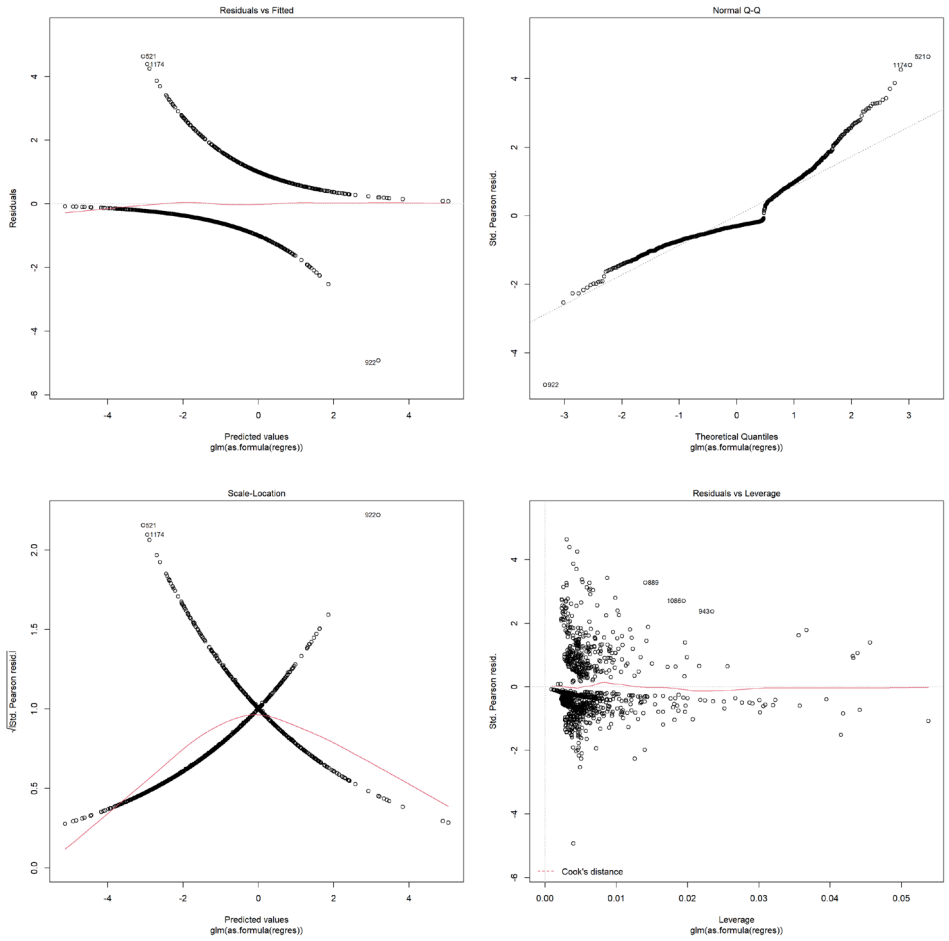


Gleason ≥4+3

Model 1 (simple) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-3.761053	0.707388	-5.317	1.06e-07 ***
Age	0.044991	0.010068	4.469	7.86e-06 ***
PSA Density (log)	0.936467	0.115292	8.123	4.56e-16 ***
Prior Biopsy	-0.868068	0.411478	-2.110	0.0349 *
MRI Volume	0.001173	0.003223	0.364	0.7160
MRI Score 4	1.144103	0.282175	4.055	5.02e-05 ***
MRI Score 5	2.143677	0.277699	7.719	1.17e-14 ***

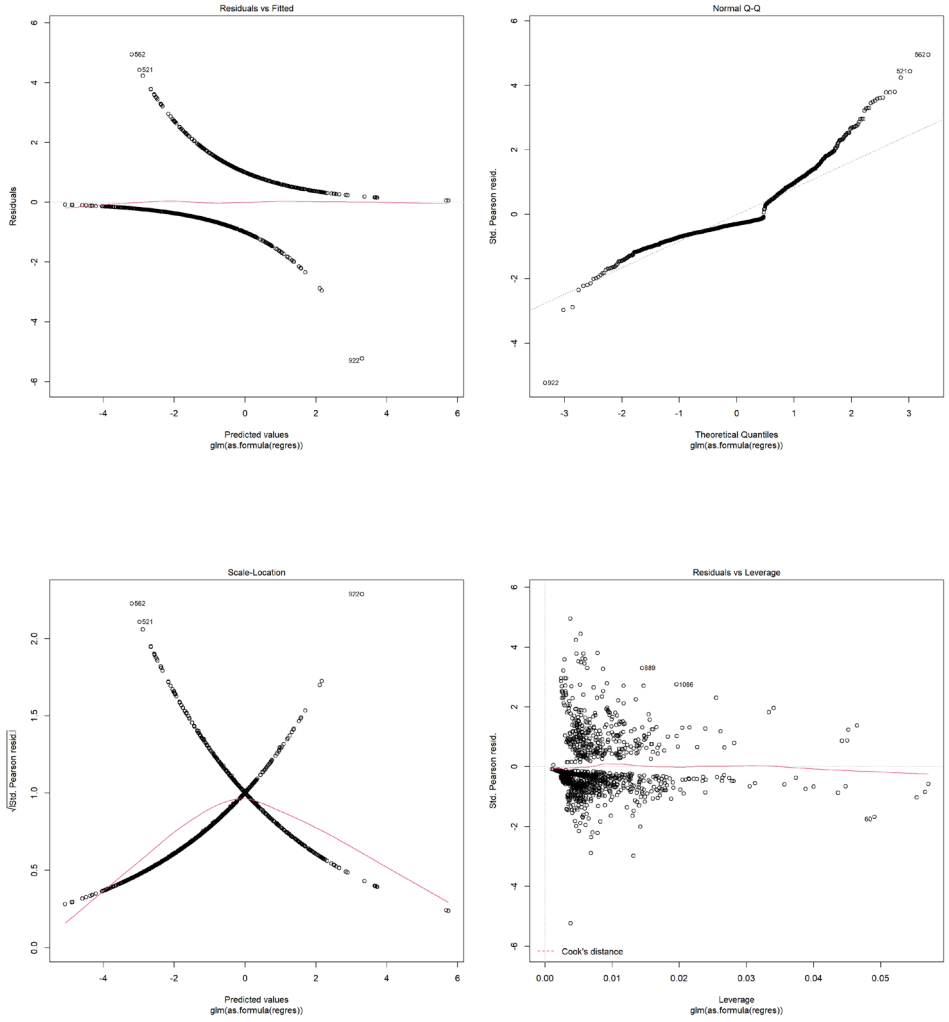
Model 1 (simple) Regression Diagnostics



Model 2 (non-invasive) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-4.5077160	0.7467421	-6.037	1.57e-09 ***
Age	0.0474892	0.0101946	4.658	3.19e-06 ***
PSA Density (log)	0.8625021	0.1198114	7.199	6.07e-13 ***
Prior Biopsy	-0.8143267	0.4167013	-1.954	0.05068 .
MRI Volume	-0.0003884	0.0033481	-0.116	0.90766
MRI Score 4	1.2895706	0.2897631	4.450	8.57e-06 ***
MRI Score 5	2.1181198	0.2786716	7.601	2.94e-14 ***
Family History	0.4250228	0.2245793	1.893	0.05842 .
Lesion_size	0.0255717	0.0089110	2.870	0.00411 **

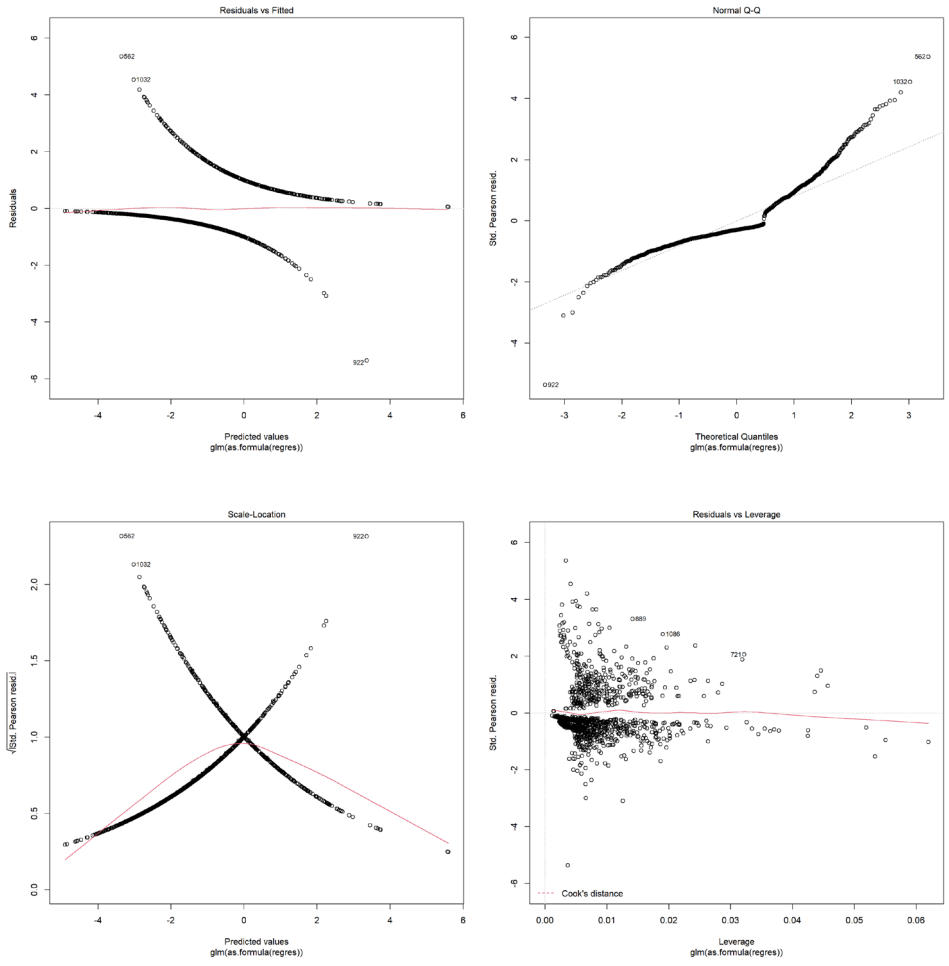
Model 2 (non-invasive) Regression Diagnostics



Model 3 (Full) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-4.5637839	0.7521562	-6.068	1.30e-09 ***
Age	0.0455857	0.0102571	4.444	8.82e-06 ***
PSA Density (log)	0.8169622	0.1204980	6.780	1.20e-11 ***
Prior Biopsy	-0.6531035	0.4170849	-1.566	0.117377
MRI Volume	-0.0005785	0.0033680	-0.172	0.863633
MRI Score 4	1.2472578	0.2896457	4.306	1.66e-05 ***
MRI Score 5	2.0207738	0.2797154	7.224	5.03e-13 ***
Family History	0.4104013	0.2266562	1.811	0.070191 .
Lesion_size	0.0217604	0.0090779	2.397	0.016527 *
DRE Abnormal	0.5427353	0.1543256	3.517	0.000437 ***

Model 3 (Full) Regression Diagnostics

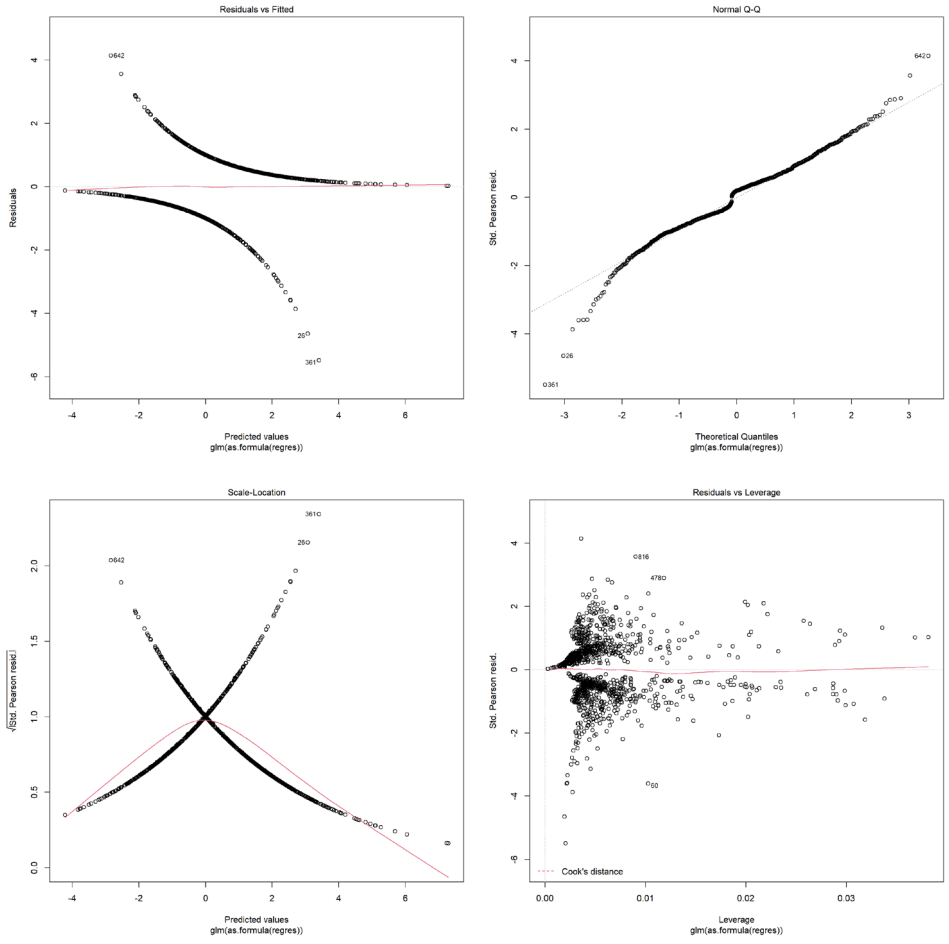


ULC/Ahmed 1

Model 1 (simple) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-1.316886	0.652738	-2.017	0.0436 *
Age	0.040704	0.009705	4.194	2.74e-05 ***
PSA Density (log)	1.116717	0.131379	8.500	< 2e-16 ***
Prior Biopsy	-0.772812	0.335485	-2.304	0.0212 *
MRI Volume	-0.007571	0.003268	-2.317	0.0205 *
MRI Score 4	0.868963	0.189278	4.591	4.41e-06 ***
MRI Score 5	2.019578	0.204008	9.899	< 2e-16 ***

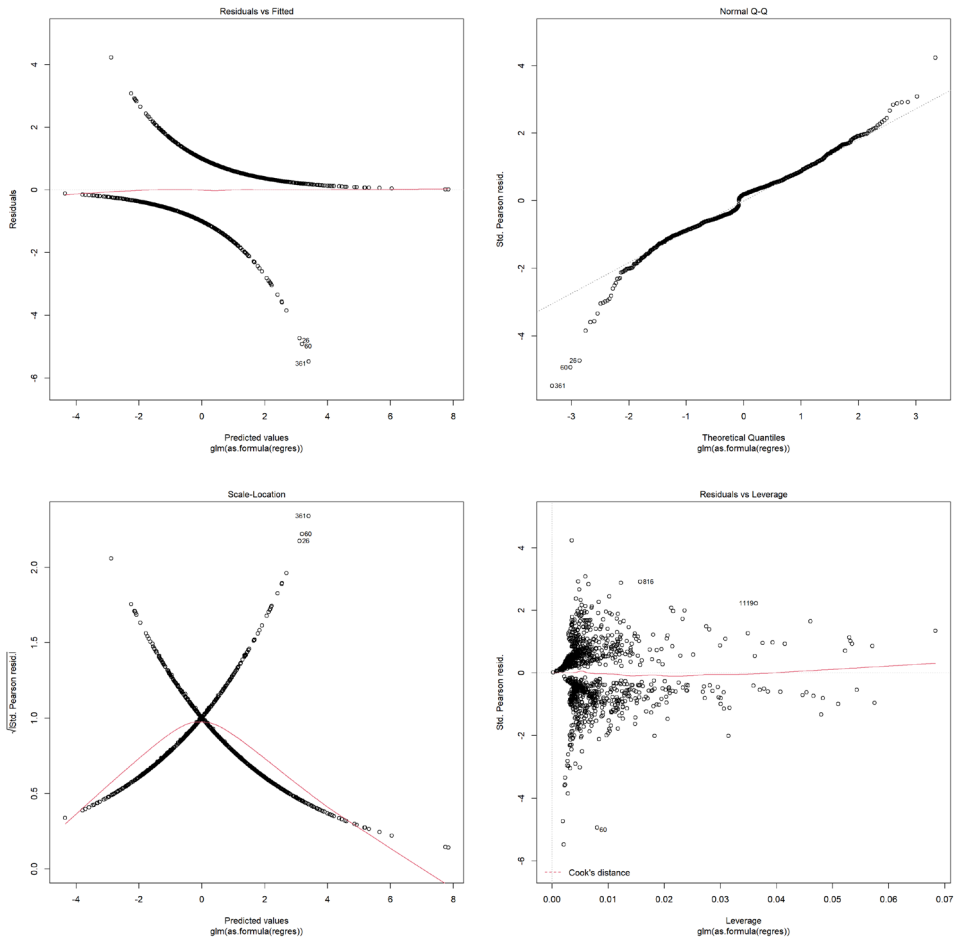
Model 1 (simple) Regression Diagnostics



Model 2 (non-invasive) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-1.712338	0.667926	-2.564	0.0104 *
Age	0.046152	0.009908	4.658	3.19e-06 ***
PSA Density (log)	1.114492	0.131862	8.452	< 2e-16 ***
Prior Biopsy	-0.702667	0.339103	-2.072	0.0383 *
MRI Volume	-0.007759	0.003275	-2.369	0.0178 *
MRI Score 4	0.861601	0.190250	4.529	5.93e-06 ***
MRI Score 5	2.047045	0.205680	9.953	< 2e-16 ***
5-ARIs	-1.078049	0.448237	-2.405	0.0162 *
Family History	0.478040	0.214099	2.233	0.0256 *

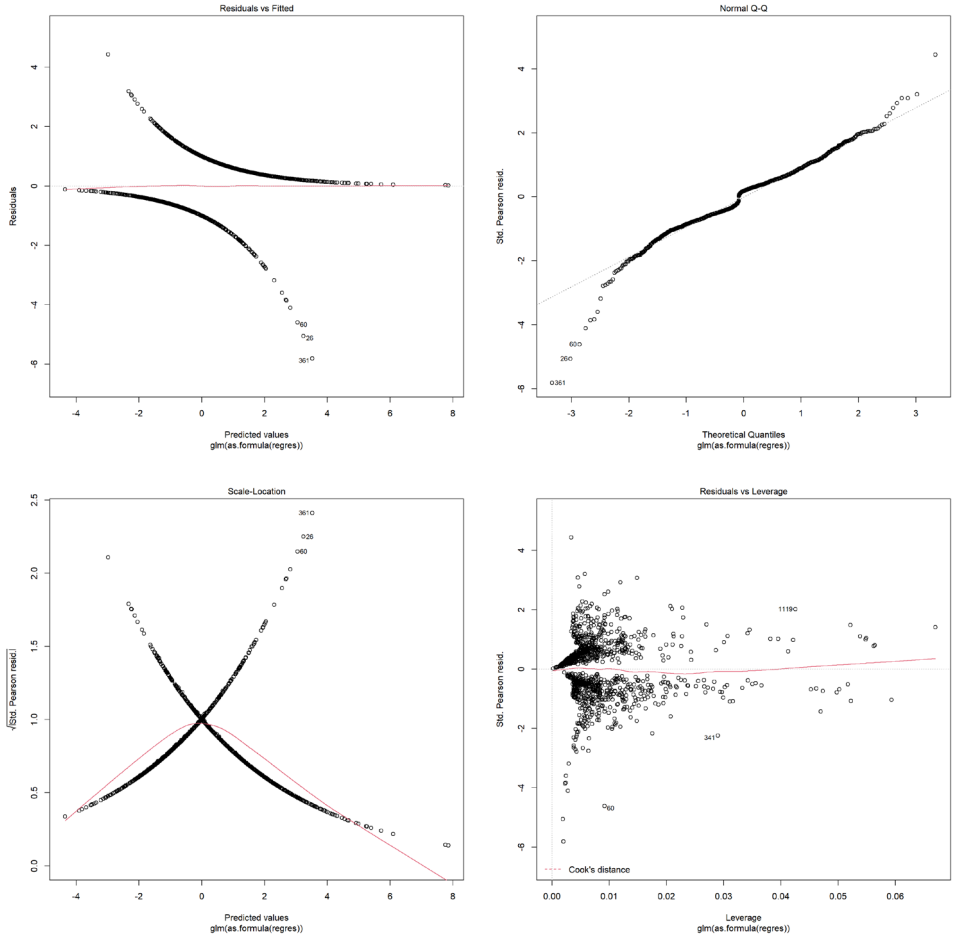
Model 2 (non-invasive) Regression Diagnostics



Model 3 (Full) Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-1.777857	0.672159	-2.645	0.00817 **
Age	0.044804	0.009968	4.495	6.97e-06 ***
PSA Density (log)	1.085658	0.132396	8.200	2.40e-16 ***
Prior Biopsy	-0.606826	0.339745	-1.786	0.07408 .
MRI Volume	-0.008012	0.003291	-2.434	0.01492 *
MRI Score 4	0.854098	0.190677	4.479	7.49e-06 ***
MRI Score 5	1.979554	0.206709	9.577	< 2e-16 ***
5-ARIs	-1.109782	0.447913	-2.478	0.01322 *
Family History	0.479017	0.214911	2.229	0.02582 *
DRE Abnormal	0.371373	0.151087	2.458	0.01397 *

Model 3 (Full) Regression Diagnostics

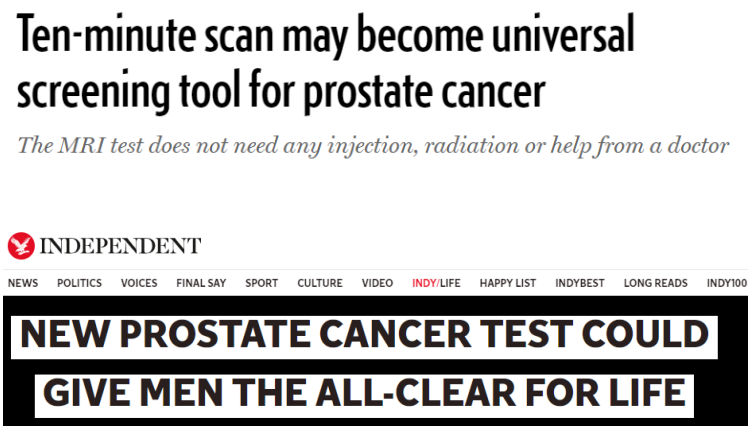


Appendix V: Publications and Proposals

Nature Reviews Urology Proposal Template

Working title:
Rethinking prostate cancer screening: Could MRI be an alternative screening test?

Rationale
We are proposing a perspective article appraising the evidence for MRI as a new screening test for prostate cancer. This is in response to widespread reports in the national media in June 2019 that all men could be offered an MRI scan as a universal screening tool for prostate cancer analogous to mammography for breast cancer or low-dose CT for lung cancer


Ten-minute scan may become universal screening tool for prostate cancer

The MRI test does not need any injection, radiation or help from a doctor

INDEPENDENT

NEWS POLITICS VOICES FINAL SAY SPORT CULTURE VIDEO INDY/LIFE HAPPY LIST INDYBEST LONG READS INDY100

NEW PROSTATE CANCER TEST COULD GIVE MEN THE ALL-CLEAR FOR LIFE

A number of clinical trials have started assessing a fast biparametic MRI as an alternative to PSA as a screening test. In this article, we provide a balanced perspective which explores the inherent challenges and potential advantages of MRI screening for prostate cancer.

The article will start by discussing the characteristics of an ideal screening test for prostate cancer based on the criteria defined by Wilson and Jungner. We will closely examine whether PSA meets these attributes using the most recent evidence from large-scale randomised controlled trials on PSA screening.

The second part will evaluate MRI against the same criteria. It will discuss the potential barriers and technical challenges to implementing MRI screening and we explore practical solutions including an abbreviated MRI protocol, a revised scoring system and targeted screening of high-risk groups. The article will conclude by highlighting that there are lessons to be learnt from the widespread introduction of PSA screening and describing a stepwise research framework to ensure the proper evaluation of the role of MRI as a screening test.

Eldred-Evans, David, Paula Burak, Martin J. Connor, Emily Day, Martin Evans, Francesca Fiorentino, Martin Gammon, et al. 'Population-Based Prostate Cancer Screening With Magnetic Resonance Imaging or Ultrasonography: The IP1-PROSTAGRAM Study'. *JAMA Oncology* 7, no. 3 (1 March 2021): 395. https://doi.org/10.1001/jamaoncol.2020.7456.

Eldred-Evans, David, Joana B. Neves, Lucy A.M. Simmons, Abi Kanthabalan, Neil McCartan, Taimur T. Shah, Manit Arya, et al. 'Added Value of Diffusion-Weighted Images and Dynamic Contrast Enhancement in Multiparametric Magnetic Resonance Imaging for the Detection of Clinically Significant Prostate Cancer in the PICTURE Trial: Added Value of Functional mpMRI Sequences in PICTURE'. *BJU International* 125, no. 3 (March 2020): 391–98. https://doi.org/10.1111/bju.14953.

Eldred-Evans, David, Henry Tam, Heminder Sokhi, Anwar R. Padhani, Mathias Winkler, and Hashim U. Ahmed. 'Rethinking Prostate Cancer Screening: Could MRI Be an Alternative Screening Test?' *Nature Reviews Urology* 17, no. 9 (September 2020): 526–39. https://doi.org/10.1038/s41585-020-0356-2.