

Development, validation and evaluation of an instrument for active monitoring of men with clinically localised prostate cancer: systematic review, cohort studies and qualitative study

Andrew J Simpkin, Leila Rooshenas, Julia Wade, Jenny L Donovan, J Athene Lane, Richard M Martin, Chris Metcalfe, Peter C Albertsen, Freddie C Hamdy, Lars Holmberg, David E Neal and Kate Tilling



Development, validation and evaluation of an instrument for active monitoring of men with clinically localised prostate cancer: systematic review, cohort studies and qualitative study

Andrew J Simpkin,¹ Leila Rooshenas,¹ Julia Wade,¹ Jenny L Donovan,¹ J Athene Lane,¹ Richard M Martin,¹ Chris Metcalfe,¹ Peter C Albertsen,² Freddie C Hamdy,³ Lars Holmberg,^{4,5} David E Neal⁶ and Kate Tilling^{1*}

¹School of Social and Community Medicine, University of Bristol, Bristol, UK

²Division of Urology, University of Connecticut Health Center, Farmington, CT, USA

³Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

⁴Faculty of Life Sciences & Medicine, King's College London, London, UK

⁵Regional Cancer Centre, Uppsala/Örebro Region, Uppsala, Sweden

⁶Department of Oncology, University of Cambridge, Cambridge, UK

*Corresponding author

Declared competing interests of authors: none

Published July 2015

DOI: 10.3310/hsdr03300

This report should be referenced as follows:

Simpkin AJ, Rooshenas L, Wade J, Donovan JL, Lane JA, Martin RM, *et al.* Development, validation and evaluation of an instrument for active monitoring of men with clinically localised prostate cancer: systematic review, cohort studies and qualitative study. *Health Serv Deliv Res* 2015;**3**(30).

Health Services and Delivery Research

ISSN 2050-4349 (Print)

ISSN 2050-4357 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HS&DR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Services and Delivery Research* journal

Reports are published in *Health Services and Delivery Research* (HS&DR) if (1) they have resulted from work for the HS&DR programme or programmes which preceded the HS&DR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

HS&DR programme

The Health Services and Delivery Research (HS&DR) programme, part of the National Institute for Health Research (NIHR), was established to fund a broad range of research. It combines the strengths and contributions of two previous NIHR research programmes: the Health Services Research (HSR) programme and the Service Delivery and Organisation (SDO) programme, which were merged in January 2012.

The HS&DR programme aims to produce rigorous and relevant evidence on the quality, access and organisation of health services including costs and outcomes, as well as research on implementation. The programme will enhance the strategic focus on research that matters to the NHS and is keen to support ambitious evaluative research to improve health services.

For more information about the HS&DR programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hsdr>

This report

The research reported in this issue of the journal was funded by the HS&DR programme or one of its preceding programmes as project number 09/2000/63. The contractual start date was in July 2011. The final report began editorial review in January 2014 and was accepted for publication in December 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2015. This work was produced by Simpkin *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Health Services and Delivery Research Editor-in-Chief

Professor Ray Fitzpatrick Professor of Public Health and Primary Care, University of Oxford, UK

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

Professor John Norrie Health Services Research Unit, University of Aberdeen, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

Development, validation and evaluation of an instrument for active monitoring of men with clinically localised prostate cancer: systematic review, cohort studies and qualitative study

Andrew J Simpkin,¹ Leila Rooshenas,¹ Julia Wade,¹ Jenny L Donovan,¹ J Athene Lane,¹ Richard M Martin,¹ Chris Metcalfe,¹ Peter C Albertsen,² Freddie C Hamdy,³ Lars Holmberg,^{4,5} David E Neal⁶ and Kate Tilling^{1*}

¹School of Social and Community Medicine, University of Bristol, Bristol, UK

²Division of Urology, University of Connecticut Health Center, Farmington, CT, USA

³Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

⁴Faculty of Life Sciences & Medicine, King's College London, London, UK

⁵Regional Cancer Centre, Uppsala/Örebro Region, Uppsala, Sweden

⁶Department of Oncology, University of Cambridge, Cambridge, UK

*Corresponding author kate.tilling@bristol.ac.uk

Background: Active surveillance [(AS), sometimes called active monitoring (AM)], is a National Institute for Health and Care Excellence-recommended management option for men with clinically localised prostate cancer (PCa). It aims to target radical treatment only to those who would benefit most. Little consensus exists nationally or internationally about safe and effective protocols for AM/AS or triggers that indicate if or when men should move to radical treatment.

Objective: The aims of this project were to review how prostate-specific antigen (PSA) has been used in AM/AS programmes; to develop and test the validity of a new model for predicting future PSA levels; to develop an instrument, based on PSA, that would be acceptable and effective for men and clinicians to use in clinical practice; and to design a robust study to evaluate the cost-effectiveness of the instrument.

Methods: A systematic review was conducted to investigate how PSA is currently used to monitor men in worldwide AM/AS studies. A model for PSA change with age was developed using Prostate testing for cancer and Treatment (ProtecT) data and validated using data from two PSA-era cohorts and two pre-PSA-era cohorts. The model was used to derive 95% PSA reference ranges (PSARRs) across ages. These reference ranges were used to predict the onset of metastases or death from PCa in one of the pre-PSA-era cohorts. PSARRs were incorporated into an active monitoring system (AMS) and demonstrated to 18 clinicians and 20 men with PCa from four NHS trusts. Qualitative interviews investigated patients' and clinicians' views about current AM/AS protocols and the acceptability of the AMS within current practice.

Results: The systematic review found that the most commonly used triggers for clinical review of PCa were PSA doubling time (PSADT) < 3 years or PSA velocity (PSAv) > 1 ng/ml/year. The model for PSA change (developed using ProtecT study data) predicted PSA values in AM/AS cohorts within 2 ng/ml of observed PSA in up to 79% of men. Comparing the three PSA markers, there was no clear optimal approach to alerting men to worsening cancer. The PSARR and PSADT markers improved the model c-statistic for predicting death from PCa by 0.11 (21%) and 0.13 (25%), respectively, compared with using

diagnostic information alone [PSA, age, tumour stage (T-stage)]. Interviews revealed variation in clinical practice regarding eligibility and follow-up protocols. Patients and clinicians perceive current AM/AS practice to be framed by uncertainty, ranging from uncertainty about selection of eligible AM/AS candidates to uncertainty about optimum follow-up protocols and thresholds for clinical review/radical treatment. Patients and clinicians generally responded positively to the AMS. The impact of the AMS on clinicians' decision-making was limited by a lack of data linking AMS values to long-term outcomes and by current clinical practice, which viewed PSA measures as one of several tools guiding clinical decisions in AM/AS. Patients reported that they would look to clinicians, rather than to a tool, to direct decision-making.

Limitations: The quantitative findings were severely hampered by a lack of clinical outcomes or events (such as metastases). The qualitative findings were limited through reliance on participants' reports of practices and recollections of events rather than observations of actual interactions.

Conclusions: Patients and clinicians found that the instrument provided additional, potentially helpful, information but were uncertain about the current usefulness of the risk model we developed for routine management. Comparison of the model with other monitoring strategies will require clinical outcomes from ongoing AM/AS studies.

Funding: The National Institute for Health Research Health Services and Delivery Research programme.

Contents

List of tables	xi
List of figures	xiii
List of boxes	xv
Glossary	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Aim	1
Objectives	1
Background	1
<i>Prostate cancer</i>	1
<i>Screening for prostate cancer</i>	1
<i>Treatment for clinically localised prostate cancer</i>	1
<i>Active monitoring or surveillance</i>	2
<i>Prostate-specific antigen kinetics</i>	2
<i>Modelling prostate-specific antigen change with age</i>	3
<i>Qualitative research in active monitoring/active surveillance</i>	3
<i>NHS policy and practice</i>	4
Overview of this report	5
Chapter 2 Systematic review of active monitoring for prostate cancer	7
Aim	7
Search strategy	7
<i>Inclusion criteria</i>	7
<i>Data extraction</i>	7
<i>Data synthesis</i>	7
<i>Bias in reporting</i>	8
Results	8
<i>Eligibility criteria</i>	9
<i>Surveillance or monitoring protocol</i>	15
<i>Triggers for clinical review or radical treatment</i>	15
<i>Radical treatment rates</i>	15
<i>Study factors associated with the rate of management change</i>	15
<i>Reasons for changing management</i>	23
<i>Prostate-cancer-specific mortality and metastases</i>	23
Summary	23

Chapter 3 Building a model for prostate-specific antigen during active monitoring	25
Aims and background	25
Methods	25
<i>Study population</i>	25
<i>Methods for modelling repeated measures of prostate-specific antigen</i>	26
<i>Prostate-specific antigen reference ranges</i>	26
<i>Conditional predictions for multilevel models</i>	27
Results	27
<i>Linear mixed model</i>	27
<i>Fractional polynomial mixed model</i>	28
<i>Regression spline mixed model</i>	30
<i>Functional principal components analysis</i>	30
Choosing a model for prostate-specific antigen reference ranges	32
<i>Describing the mean pattern of prostate-specific antigen change</i>	32
<i>Predicting prostate-specific antigen for a future patient</i>	33
Summary	33
Chapter 4 Testing the accuracy of the prostate-specific antigen model in four separate cohorts	35
Aims and background	35
Methods	35
<i>Study populations</i>	35
<i>Measuring performance</i>	36
Results	37
<i>Age-related prostate-specific antigen change</i>	37
<i>Screen-detected cohorts</i>	38
<i>Clinically detected cohorts</i>	39
<i>Comparison across cohorts</i>	39
Summary	40
Chapter 5 Comparison of prostate-specific antigen markers used to alert men to worsening prostate cancer	41
Aims and background	41
Methods	41
<i>Prostate-specific antigen markers</i>	41
<i>Study populations</i>	42
<i>Outcomes</i>	42
<i>Statistical analysis</i>	43
Results	43
<i>Metastases</i>	43
<i>Prostate cancer-specific mortality</i>	44
<i>Reaching a prostate-specific antigen of 20 ng/ml during surveillance</i>	45
Discussion	46
Chapter 6 Qualitative study: clinicians' and men's views on current active surveillance practices	47
Introduction	47
Methods	47
<i>Design</i>	47
<i>Sampling and recruitment of clinicians/nurses</i>	48
<i>Sampling and recruitment of men on active surveillance</i>	48
<i>Data-collection methods</i>	49
<i>Analysis</i>	49

Results	50
<i>Participant characteristics</i>	50
<i>Presentation of data</i>	50
<i>Overview</i>	51
<i>Deciding whether or not to follow an active surveillance programme</i>	51
<i>Current practice within active surveillance programmes: follow-up procedures and triggers for radical treatment</i>	54
Conclusion	66
Chapter 7 Qualitative study: clinicians' and men's views on the active monitoring system	69
Introduction	69
Methods	70
Results	70
<i>Clinicians' views on data presented in the active monitoring system</i>	70
<i>Men's perspectives on data presented in the active monitoring system</i>	75
<i>Clinicians' and men's views on presentation of data within the active monitoring system</i>	79
Conclusion	81
Chapter 8 Randomised controlled trials of active monitoring strategies	83
Aim and background	83
Difficulties related to new randomised controlled trials of active monitoring strategies	83
<i>Recruitment and design</i>	83
<i>Eligibility</i>	84
<i>Outcomes</i>	84
Potential trial designs	84
<i>Active monitoring system</i>	84
<i>Active monitoring compared with active surveillance</i>	85
<i>Active monitoring for higher-risk or locally advanced prostate cancer</i>	85
<i>Nurse-led active monitoring</i>	85
Summary	86
Chapter 9 Conclusions and recommendations	87
Summary of findings	87
Discussion	88
Research recommendations	90
<i>Modelling the relationship between prostate-specific antigen and prostate cancer outcomes</i>	90
<i>Reporting in active monitoring/active surveillance studies</i>	91
<i>Changing clinical environment</i>	91
<i>Future qualitative research in active monitoring/active surveillance</i>	91
Acknowledgements	93
References	95
Appendix 1 Development of active monitoring system	105
Appendix 2 Dissemination	107
Appendix 3 Qualitative documentation	109

Appendix 4 Description of statistical methods used to model repeated measures of prostate-specific antigen	131
Appendix 5 Description of the studies and data used for analysis	135

List of tables

TABLE 1 Eligibility criteria, monitoring protocol and characteristics for the 22 included studies, sorted by year of commencing recruitment	10
TABLE 2 Management change definitions and results for the 22 included studies	16
TABLE 3 Comparison of fixed-effect estimates and standard errors from three parametric models	27
TABLE 4 Results of the fractional polynomial testing procedure	28
TABLE 5 Fits to the observed PSA (ng/ml) by age category	32
TABLE 6 Root-mean-square error of PSA (ng/ml) by age category	32
TABLE 7 Model diagnostics and RMSE (in ng/ml) for the parametric approaches (no AIC is available from the PACE method)	33
TABLE 8 Coefficients from simple models of PSA change across six cohorts	37
TABLE 9 Average absolute difference between observed and predicted PSA and coverage of observed PSA using the protect model on four separate cohorts	38
TABLE 10 Model failure by study using prediction conditioned on first, second and third PSA	39
TABLE 11 Model success by study using prediction conditioned on first, second and third PSA	39
TABLE 12 Summary of PCa outcomes used for validation in each cohort, with the percentage positive for each outcome during cohort follow-up	43
TABLE 13 Sensitivity and specificity comparing the PSA markers in alerting men with onset of metastases	44
TABLE 14 Sensitivity and specificity comparing the PSA markers in alerting men with PCa specific mortality	44
TABLE 15 Hazard ratio and c-statistic improvements for the three PSA markers on each of the four outcomes	45
TABLE 16 Differences in deviance (predictive ability) between the three PSA marker models	45
TABLE 17 Characteristics of men interviewed	50
TABLE 18 Clinicians' reported preferred maximum thresholds for clinical parameters considered in decision-making surrounding AS eligibility	52
TABLE 19 Decision-making leading to initiation of AS programme	53

TABLE 20 Mr Smith paper vignette	114
TABLE 21 Mr Jones paper vignette	114
TABLE 22 Mr Evans paper vignette	115
TABLE 23 Quality assurance checklist for reporting of qualitative research	127
TABLE 24 Description of the studies used for analysis	136
TABLE 25 Characteristics of the data received	137

List of figures

FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart of the screening process	9
FIGURE 2 Forest plot of the proportion of men changing to radical treatment per person-years, sorted by year of beginning recruitment	22
FIGURE 3 Plot of PSA trajectories for five simulated men	26
FIGURE 4 Fits to the five simulated PSA profiles	28
FIGURE 5 Mean PSA over age (solid line) with fits using LMM (dashed), FPMM (dotted), RSMM (dash dotted) and PACE (long dashed)	31
FIGURE 6 Hypothetical PSA change in the cohorts if average PSA at 50 years was 2 ng/ml in each	37
FIGURE 7 Estimated PSA change in the cohorts	38
FIGURE 8 An example of a PSARR	42
FIGURE 9 Active monitoring system	69
FIGURE 10 The PSARR system for a simulated individual, showing average PSA trend (blue, short dash) plus 95% reference range (orange) for men without PCa, as well as average PSA trend for men with PCa (green, long dash)	105
FIGURE 11 Mr Smith vignette presented on AMS	116
FIGURE 12 Mr Jones vignette presented on AMS	117
FIGURE 13 Mr Evans vignette presented on AMS	117
FIGURE 14 Fitting process of functional PCA	133

List of boxes

BOX 1 Inclusion and exclusion criteria for selecting men	48
BOX 2 Clinicians' interview topic guide	109
BOX 3 Men's interview topic guide	112

Glossary

Active monitoring A management option for men with clinically localised prostate cancer. Active monitoring requires pre-defined eligibility criteria, a management protocol and triggers for clinical review leading to change of management. Men are followed up with repeated prostate-specific antigen testing and, often, digital rectal examinations, with the option for rebiopsy if the latter appear to show worsening disease. Regular review provides the opportunity to consider all management options, including radical treatment, on request or where worsening results from prostate-specific antigen testing and digital rectal examinations are evident.

Active monitoring system A Microsoft Excel® 2013 (Microsoft Corporation, Redmond, WA, USA) spreadsheet that records and displays repeated measurements of prostate-specific antigen in a graphical format.

Active surveillance Similar to active monitoring but usually with protocols that allow entry of only men with defined low-risk prostate cancer. Active surveillance also includes scheduled repeat biopsy testing of men, as well as prostate-specific antigen and digital rectal examination testing.

Cox proportional hazards model A regression type model for time-to-event data, such as time to onset of metastases.

Fractional polynomial mixed model A statistical method by which smooth curves can be fitted to repeated measures of, for example, prostate-specific antigen over time.

Functional principal components analysis A statistical method by which smooth curves can be fitted to repeated measures of, for example, prostate-specific antigen over time.

Linear mixed model A statistical method by which lines can be fitted to repeated measures of, for example, prostate-specific antigen over time.

Principal components analysis through conditional expectation A statistical method by which smooth curves can be fitted to repeated measures of, for example, prostate-specific antigen over time.

Prostate-specific antigen A serum that is produced in the prostate. Prostate-specific antigen is a key biomarker in the diagnosis and monitoring of prostate cancer.

Prostate-specific antigen doubling time A summary measure of repeated measurements of prostate-specific antigen which corresponds to the estimated time it will take for a man's prostate-specific antigen to double. A low positive value of prostate-specific antigen doubling time corresponds to rapidly rising prostate-specific antigen.

Prostate-specific antigen reference ranges A model-based approach to alerting men to rapidly rising prostate-specific antigen during active monitoring.

Prostate-specific antigen velocity A summary measure of repeated measurements of prostate-specific antigen which corresponds to the rate of change of prostate-specific antigen. A high value of prostate-specific antigen velocity corresponds to rapidly rising prostate-specific antigen.

Prostate testing for cancer and Treatment trial An ongoing multicentre randomised controlled trial comparing active monitoring, radical prostatectomy and radiotherapy in the UK.

Regression spline mixed model A statistical method by which smooth curves can be fitted to repeated measures of, for example, prostate-specific antigen over time.

Root-mean-square error A summary measure for comparing statistical models, which describes the average absolute difference between observed and predicted values (e.g. between observed prostate-specific antigen values and those predicted by the model).

Watchful waiting A conservative management approach for generally older men with comorbidities or other reasons for which radical treatment may not be suitable. No trigger for radical treatment is in place, although palliative care through hormonal treatment may be recommended.

List of abbreviations

AIC	Akaike information criterion	PCSM	prostate cancer-specific mortality
AM	active monitoring	PIVOT	Prostate cancer Intervention Versus Observation Trial
AMS	active monitoring system	Protect	Prostate testing for cancer and Treatment trial
AS	active surveillance	PSA	prostate-specific antigen
AUC	area under the curve	PSAD	prostate-specific antigen density
BLSA	Baltimore Longitudinal Study of Aging	PSADT	prostate-specific antigen doubling time
CI	confidence interval	PSARR	prostate-specific antigen reference range
Cox PH	Cox proportional hazards model	PSAv	prostate-specific antigen velocity
DRE	digital rectal examination	RCT	randomised controlled trial
FPCA	functional principal components analysis	RMH	Royal Marsden Hospital
FPMM	fractional polynomial mixed model	RMSE	root-mean-square error
GP	general practitioner	RSMM	regression spline mixed model
HR	hazard ratio	SPCG4	Scandinavian Prostate Cancer Group study number 4
JH	Johns Hopkins	SD	standard deviation
LMM	linear mixed model	T-stage	tumour stage
MRI	magnetic resonance imaging	UCHC	University of Connecticut Health Center
NICE	National Institute for Health and Care Excellence	WW	watchful waiting
PACE	principal components analysis through conditional expectation		
PCa	prostate cancer		

Plain English summary

Active monitoring (AM) or active surveillance (AS) are management options for men with low-risk prostate cancer (PCa), consisting of regular blood tests measuring the level of prostate-specific antigen (PSA) and, sometimes, regular prostate biopsies. A rising PSA level can indicate worsening PCa, but PSA can also be raised for non-cancer-related reasons. A UK study [Prostate testing for cancer and Treatment (ProtecT)] will report results in 2016 that will provide clearer evidence about the best treatment (radiotherapy, AM or surgery) for low-risk PCa. Many research groups monitor PCa using PSA, but each uses different methods and there is much debate about the best method to use.

We developed a model for changes in PSA with age using data from men on AM in ProtecT. This model predicted PSA reasonably accurately in similar groups of men in the UK and USA. We developed software that shows the observed PSA data for a given patient in comparison with an average man's change in PSA and highlights an observed PSA in the predicted top 5%.

We interviewed patients and clinicians about current methods of AM/AS and the acceptability of our software. Patients and clinicians expressed reservations about whether or not the software would currently influence decision-making. Clinicians did not feel that they would make recommendations about stopping or continuing AM/AS based on PSA monitoring alone. Patients maintained that decision-making would ultimately be guided by their clinician.

Scientific summary

Background

Prostate cancer (PCa) is one of the most common cancers and one of the most common causes of cancer death in the UK. Men diagnosed with clinically localised PCa can choose between radical treatment (prostatectomy, radiotherapy) or active surveillance (AS), the latter of which aims to target radical treatment only to those who would benefit most. AS consists of regular check-ups of prostate-specific antigen (PSA; an antigen in serum measured through a blood test), digital rectal examination (DRE) and repeat biopsy. Also known as active monitoring (AM), AS triggers a clinical review with the opportunity for a biopsy when results from PSA and DRE tests appear to show worsening disease. AM/AS are distinguished from watchful waiting (WW), which is a method of managing men with prostate cancer who are not suitable for radical treatment, involving treatment only if and when they develop symptoms. The Prostate testing for cancer and Treatment (ProtecT) trial, funded by the National Institute for Health Research Health Technology Assessment programme, is the first randomised controlled trial (RCT) comparing AM with radical treatments (surgery and radiotherapy) and will report primary outcomes in 2016.

There is uncertainty over which men are suitable for AM/AS, how to monitor them and if or when to recommend radical treatment. Several ongoing studies of AM/AS use PSA kinetic measures such as PSA doubling time (PSADT) or PSA velocity (PSAv) to trigger clinical review. To use PSA to monitor men, it is essential to be able to distinguish age-related from pathological changes and to allow for the wide within-individual variation in PSA. A model for PSA changes over time, if shown to be accurate in predicting PSA across different populations of men with PCa, could be used to provide comparison values for individual PSA changes.

There has been little qualitative research into men's and clinicians' experiences and views about AM/AS. Research has tended to focus on men's treatment decision-making following diagnosis of PCa, with a smaller number of qualitative studies drawing attention to the key issue of men's uncertainty and how this is dealt with. There is little qualitative evidence about experiences of undergoing or providing AM/AS or how AM/AS is conducted in routine practice.

Objectives

1. To compare protocols and triggers for radical treatment or clinical review in studies of AM/AS worldwide.
2. To develop a model for age-related PSA change in men on AM/AS.
3. To validate this model in predicting PSA in external cohorts of men on AM/AS.
4. To use the model to derive 95% reference ranges for PSA with age.
5. To test the ability of an observed PSA value lying outside these reference ranges to predict clinical outcomes.
6. To explore the acceptability of a graphical presentation of the reference ranges [the active monitoring system (AMS)] in clinical practice through interviews with men on AM/AS and their clinicians.

Data sources

Prostate-specific antigen data were available from men diagnosed through PSA testing who had refused randomisation but had chosen to be managed by AM in the ProtecT study. These data included 7438 PSA tests from 512 men over a mean follow-up of 4.8 years [standard deviation (SD) 2.4 years], along with other clinical, social and demographic variables. Men were eligible for AM if they had a diagnosis of clinically localised PCa (T1/T2) and were also eligible for surgery or radiotherapy. They were followed up with PSA tests every 3 months in year 1 and 6-monthly thereafter, with a formal assessment of PSA change over a 12-month period each year. Given that the ProtecT trial outcome data are under embargo until publication of results in 2016, the clinical outcomes of these 512 men remain unknown.

The Royal Marsden Hospital (RMH) AS cohort data consisted of 499 men and 9427 PSA test results over an average of 4.5 years (SD 2.6 years). Clinical outcomes of metastases and PCa-specific mortality (PCSM) were available alongside baseline measurements of Gleason score, tumour stage (T-stage), percentage of free PSA, prostate volume and percentage of positive biopsy cores. These data were collected between 1999 and 2012 and represent a modern cohort of UK men on AS, with the majority diagnosed through a raised PSA level and subsequent positive biopsy. The study had eligibility criteria of baseline PSA < 15 ng/ml, Gleason score of $\leq 3 + 4$ and percentage of positive biopsy cores $\leq 50\%$. Men on AS were followed up with PSA tests every 3–4 months in the first 2 years and every 6 months thereafter. Biochemical progression was defined as PSA_v > 1 ng/ml/year, whereas histological progression on rebiopsy was Gleason score of $\geq 4 + 3$ or > 50% positive biopsy cores.

A similar US cohort is the Johns Hopkins (JH) Hospital AS study. We received PSA data on 961 men, comprising 9993 PSA test results, along with baseline Gleason score, prostate volume, PSA density (PSAD), percentage of free PSA, percentage of positive cores and the percentage of cancer per positive core. Clinical outcomes of all-cause and PCSM were also included. These PSA data were collected between 1992 and 2012 and represent a modern, ongoing AS population. Men were eligible if they had a Gleason score of $\leq 3 + 3$, stage T1c, PSAD < 0.15 ng/ml/cm³, ≤ 2 positive biopsy cores and maximum involvement of 50% per core. Radical treatment was recommended when men no longer met the eligibility criteria described above.

The University of Connecticut Health Center (UCHC) cohort represents an older era of WW/AM before the onset of PSA screening. The men were, for the most part, diagnosed clinically – through presentation with symptoms for example. The data included 114 men with 844 PSA test results over an average follow-up of 4.7 years (SD 3.9 years). Gleason score at diagnosis and an outcome of all-cause mortality were available.

Data were provided by the Scandinavian Prostate Cancer Group study number 4 (SPCG4) from the WW arm of their RCT. The men in this study were diagnosed between 1989 and 1999, the majority before PSA testing became widely available, and represent a cohort of men with clinically localised PCa at a more advanced stage. This cohort comprised 290 men, with 2987 PSA test results during an average of 6.0 years of follow-up (SD 3.8 years). Along with baseline Gleason score and T-stage, outcomes of metastases and PCSM were available.

Prostate-specific antigen testing was not widespread in the UK, Scandinavia or the USA at the time of the SPCG4 and UCHC cohorts, whereas screening is common in the USA today. The SPCG4 and UCHC cohorts thus represent clinically presenting men, not PSA-detected men. The ProtecT and RMH cohorts represent men detected by an initial PSA test at an arbitrary time-point (who could have had raised PSA for some time before inclusion in this study). The JH cohort represents men detected via ongoing and repeated screening (and who are thus unlikely to have prior raised PSA). The cohorts thus represent a wide range of ‘types’ of PCa diagnosed at different points along the long lead-time in the development of PCa.

Methods

A systematic review was undertaken to examine AM/AS studies worldwide, particularly methods of monitoring PSA and the triggers used to recommend clinical review or radical treatment. The rate of management change from AM/AS to radical treatment was estimated through meta-analysis.

A model for PSA change with age was developed using the ProtecT trial data. Multilevel modelling was used to account for variability in PSA values between and within individuals over the course of the monitoring period. The resulting ProtecT model was externally validated in men from the RMH, JH, SPCG4 and UCHC cohorts.

The model was used to develop 95% reference ranges for PSA with age. Observed PSA values above the reference range would account for the top 5% of PSA level and could be indicative of pathological changes in PCa status. Thus, men with a single PSA test result above their PSA reference ranges (PSARR) would be recommended for clinical review. Specificity and sensitivity of PSARR, PSADT and PSA_v were compared, using metastases and PCSM as binary outcomes. Cox proportional hazards models were used to compare the predictive ability of the three PSA markers using these outcomes as time-to-event data. A c-statistic was calculated from a model using diagnostic (i.e. baseline) information from each individual, and improvements in the c-statistic from this model, as well as model deviance, were used to compare the PSA markers.

A qualitative study was designed to investigate acceptability to patients and clinicians of the PSARR model presented using Microsoft Excel® software (Microsoft Corporation, Redmond, WA, USA), referred to as the AMS. The qualitative study also sought to investigate patient and clinician perspectives on current AM/AS management and how the AMS would compare and fit with current practice.

Semi-structured interviews were conducted with 18 clinicians and 20 patients from four NHS acute trusts. Sites were purposefully selected on the basis of having a fully running AM/AS programme. Clinicians were purposefully sampled on the basis of having responsibility for leading AM/AS clinics. Clinicians identified low-risk patients for interview based on clinical eligibility criteria provided by the research team and distributed recruitment packs to potential patient participants.

Interviews took place in private, confidential settings within hospital premises or patients' homes between April and November 2013. Interviews were recorded following receipt of written consent and transcribed in full. Thematic analysis, guided by the constant comparison method from grounded theory, took place concurrently with data collection. Data collection proceeded until data saturation was achieved – after four consecutive interviews produced no new themes.

Results

The systematic review found little consensus on the optimal design of an AM/AS protocol. AS programmes were more common than AM, with 16 AS and six AM studies worldwide. PSA testing was used by all 22 studies, with various frequencies of testing. Twelve cohorts used PSA or PSA kinetics to recommend clinical review, with the most common thresholds being a PSADT < 3 years or a PSA_v > 1 ng/ml/year. The overall rate of management change was 84 per 1000 person-years [95% confidence interval (CI) 61 to 106 per 1000 person-years], with wide variation between studies. There were eight deaths from PCa in 7111 men followed for a median of 3.7 years.

Our analysis of the ProtecT trial data found that the optimal model for PSA change was a regression spline mixed model (RSMM). This model allows for a different linear change in log-PSA at different ages during AM/AS. It was found to have the best accuracy to fit the repeated PSA data from the ProtecT trial, with an average absolute difference between observed and predicted PSA of 2.1 ng/ml.

The ProtecT model predicted PSA values from the RMH and JH cohorts, with average absolute differences between predicted and observed PSA of 2.0 ng/ml in the RMH cohort and 1.8 ng/ml in the JH cohort. Between 67% and 79% of men in these two cohorts had an average difference of < 2 ng/ml per PSA test. In the older cohorts, as expected, the ProtecT model performed less well, with an average absolute difference per PSA test of 4.6 ng/ml in the SPCG4 cohort and 3.7 ng/ml in the UCHC cohort. Just 39% of men in SPCG4 and 51% of men in UCHC had an average absolute difference of < 2 ng/ml per PSA test.

Prostate-specific antigen reference ranges had the highest specificity (correctly not alerting men with stable cancer) but lowest sensitivity (correctly identifying men with lethal cancer) among the three PSA markers in both predicting metastases and death from PCa in SPCG4. In men with a Gleason score of 6, there was some weak evidence that PSADT improved on using PSA_v for metastases ($p = 0.06$) and PCSM ($p = 0.05$). In the SPCG4 cohort, a model using age, T-stage and PSA at diagnosis was not very useful for either outcome (c -statistic = 0.58 for metastases and 0.53 for PCSM). For PCSM, PSARRs (c -statistic increase 0.11 = 21%) and PSADT (c -statistic increase 0.13 = 25%) were both useful in improving discriminative ability, compared with using diagnostic information alone (PSA, age, T-stage). In the RMH and UCHC cohorts, a lack of events in clinical outcomes did not allow us to compare the PSA markers.

Interviews with clinicians revealed an appreciation of the many uncertainties underpinning current AM/AS practice. This uncertainty was reflected in variation in clinical practice in relation to (1) decisions about patient eligibility; (2) the frequency and nature of follow-up; and (3) thresholds for moving from AM/AS to radical treatment. Clinicians questioned the value of the AMS given the current lack of data linking the PSARR to clinical outcomes, and current practice where PSA kinetics are viewed as one of several tools influencing decisions about ongoing AM/AS. However, some clinicians did recognise the AMS's potential for reassurance for some men.

Patient interview participants also showed awareness of the uncertainties in the evidence base for PSA monitoring. Despite this, patients reported feeling at ease and appeared confident with their AM/AS strategy. Patients trusted clinicians to direct decision-making, including the point at which radical treatment should be considered. On this basis, patients did not feel the AMS would influence their decision-making, although they did think that some of the material presented could be helpful.

Conclusions

There is little consensus over eligibility and trigger criteria in AM/AS protocols and no published RCTs including AM or AS strategies. Observational evidence suggests very low rates of progression to life-threatening disease, but most cohorts have been highly selective and are still insufficiently mature to establish this definitively. As there is so little consensus on AM/AS study design, it would be helpful if investigators specified clear designs of eligibility criteria, monitoring protocol and triggers for clinical review. If these clear designs are strictly adhered to then results from ongoing studies of AM and AS will be able to be combined in future to obtain evidence about more effective protocols.

We found the optimal model for PSA allowed the linear trend of log-PSA to change during monitoring at ages 63 and 68 years. The ProtecT model for PSA change was useful in predicting future PSA in men undergoing AM/AS in the UK or USA.

There was no clear optimal PSA marker found when testing against PCSM, metastases and reaching a PSA of 20 ng/ml. Our data have suggested that PSADT and PSARR improved on PSA_v. Rigorous testing cannot be achieved without better outcome data, such as will arrive with the publication of the ProtecT trial.

Patients and clinicians perceive current AM/AS practices to be framed by uncertainty. These uncertainties extend from the selection of eligible AM/AS candidates to optimum follow-up protocols and effective thresholds for moving from AM/AS to clinical review or radical treatment. Patient and clinician uncertainty surrounding the effectiveness of PSA monitoring generally led to the perception that, although the AMS could add to clinical knowledge, its current impact on practice was limited until more robust data about AS become available.

Funding

The National Institute for Health Research Health Services and Delivery Research programme.

Chapter 1 Introduction

Aim

Our aims were to develop longitudinal age-related reference ranges that predict how prostate-specific antigen (PSA) levels change over time and to examine their use in identifying abnormal (cancer-related) PSA increases in men with clinically localised prostate cancer (PCa).

Objectives

- To use data from four cohorts of men with PCa to investigate whether or not data collected during diagnosis about the type, quantity and aggressiveness of the cancer (e.g. Gleason score, number of positive biopsy cores) improve the accuracy of our previously published reference ranges.
- To use data from three cohorts of men with PCa to (1) calibrate the thresholds of the reference ranges to identify when an increase in PSA might be indicative of progressing cancer; (2) calculate the sensitivity, specificity and predictive value of the reference ranges to predict clinical cancer progression; and (3) compare these predictive abilities to those of standard PSA kinetics.
- To develop an easy-to-use instrument based on the reference ranges that would assist in the clinical management of men with clinically localised PCa who have opted for active monitoring [(AM) regular PSA tests to monitor the cancer], and to explore the acceptability of various formats and presentations of the instrument to patients and clinicians.
- To design a randomised controlled trial (RCT) to evaluate the effectiveness of the most acceptable AM instrument compared with existing AM methods.

Background

Prostate cancer

For men in the UK, PCa accounts for 24% of cancers diagnosed and 14% of cancer mortality;¹ yet, currently, its natural history is not fully understood. Autopsy results have shown that approximately half of men over 50 years of age have some form of PCa.² However, based on data from the USA, there is a 16.5% lifetime chance of being diagnosed with, and a 3% chance of dying from, PCa.^{3,4}

Screening for prostate cancer

Prostate cancer is one of the most common newly detected cancers in men worldwide, primarily because of the increasing use of PSA as a screening test.⁵ PCa screening is, however, a controversial issue in contemporary health care.⁶ Screening using PSA can detect cancers that are still confined within the prostate⁷ when radical prostatectomy or radiotherapy could possibly achieve a cure. However, current tests cannot differentiate between tumours with biological potential for progression and the majority of slow-growing tumours, which will not cause clinical disease in a man's lifetime. The recent publication of the results of screening trials in Europe and the USA have further fuelled the controversies in this area, showing, for example, that 1410 men needed to be screened to find 48 cancers to prevent one cancer death.⁸

Treatment for clinically localised prostate cancer

Men diagnosed with clinically localised PCa are often treated by either radical prostatectomy (surgery to remove the prostate) or some form of radiotherapy. Radical treatments can cause serious side effects, particularly incontinence and impotence.⁹ There is currently a lack of robust evidence about the comparative effectiveness of treatment for PCa. A recent systematic review¹⁰ sought to compare monitoring with radical treatment but found

no comparative trials. Two randomised trials have evaluated the effectiveness of a passive strategy called watchful waiting (WW). WW historically involved palliative treatment once symptoms appeared and was aimed at older men who were not suitable for surgery or radiotherapy. The Prostate cancer Intervention Versus Observation Trial (PIVOT) recently found no difference between WW and radical prostatectomy for either all-cause mortality or PCa-specific mortality (PCSM) after at least 10 years of follow-up¹¹ for men with PSA-detected PCa. The Scandinavian Prostate Cancer Group study number 4 (SPCG4) trial¹² found that radical prostatectomy reduced PCSM compared with WW among men who had been diagnosed clinically (rather than screen detected as in PIVOT). This trial started recruitment prior to the use of PSA and, consequently, the majority of participants had clinically detected disease.¹³ The Prostate testing for cancer and Treatment (ProtecT) trial is comparing AM with radical treatment (surgery and radiotherapy) in men with PSA-detected clinically localised PCa, but the long periods of follow-up required for clinically relevant outcomes to be observed means that it will not report for at least another 2 years.^{14,15}

Interest in the safety and acceptability of monitoring for clinically localised PCa, which avoids unnecessary intervention, has grown over the past decade. Although it is accepted that AM protocols involve closer monitoring than that which is employed in the WW arm of SPCG4, recent systematic reviews have concluded that there is little evidence or expert consensus over the most effective monitoring protocol.^{10,16–18} Serial measures of PSA level are used consistently, but various aspects of PSA kinetics are used to trigger further clinical review, including PSA doubling time [(PSADT) length of time for PSA to double] and PSA velocity [(PSAv) rate of PSA change per year]. Hence, there is increasing interest in using PSA levels to monitor men with low-risk clinically localised tumours and in using strategies that would indicate when further clinical review and radical treatment would be appropriate and when men could remain monitored without intervention.¹⁹

Active monitoring or surveillance

Active monitoring and active surveillance (AS) have developed as management strategies for men with clinically localised PCa to avoid overtreatment of low-risk disease.²⁰ AS involves regular follow-up by PSA testing, digital rectal examination (DRE), review of symptoms and scheduled repeat biopsy, whereas AM only uses a rebiopsy in the presence of changing PSA or DRE findings, because a rising PSA can be indicative of worsening PCa. Under an AM protocol, men with stable PSA remain on monitoring, whereas those with rapidly rising PSA are recommended for clinical review. At clinical review, a biopsy of the prostate can be performed and, depending on the specific protocol, a patient may be recommended for or suggested to consider radical intervention (e.g. radiotherapy or prostatectomy) or continuation on AM. The key tenet of AM/AS is to balance the risk of harm of intervention in men with non-lethal cancer, with the risk of not intervening where cancer is progressing. A previous systematic review¹⁶ based on five AM/AS cohorts^{20–24} found little consensus on the eligibility criteria, optimum protocol for monitoring, the trigger(s) for radical intervention or the role of PSA kinetics in monitoring.¹⁶ Two more recent systematic reviews^{10,17} again found no consensus on eligibility criteria for AM/AS or the AM/AS regimens used.

Prostate-specific antigen kinetics

In order to use PSA in AM/AS, a model for normal PSA levels is needed as a comparator for observed PSA levels in men on AM/AS. PSADT, PSAv and absolute level of PSA are commonly used measures for monitoring a man's PSA level.¹⁸ There is little consensus on which of these to use or what threshold should be employed for each measure. There remains an absence of clinical or statistical evidence for their use in AM, and retrospective analyses have found very little association with clinical outcomes, such as metastases or PCSM.^{18,25–28} Furthermore, there are concerns about the various methods of calculation of PSADT^{29,30} and PSAv,^{31,32} as well as a great deal of variation and uncertainty about how many PSA values should be used for calculation.³³ PSA levels increase naturally with age, so a method is needed to indicate when increases in PSA are beyond normal age-related change to avoid reviews being triggered when they are not necessary.

Modelling prostate-specific antigen change with age

Modelling PSA in AM/AS has received little attention in the literature,^{33–36} despite the fact that repeated measures of PSA from a cohort of men represent a wealth of information about the behaviour of PSA. An accurate model for PSA change would allow men to observe where they lie in comparison to the general trend of PSA in similar men and whether or not any increase in their PSA is consistent with normal age-related change. Multilevel models^{37,38} allow such data to be fully exploited by modelling the variation in PSA between men in an AM cohort, as well as the variation in PSA within each individual during follow-up. We have previously used multilevel models to model PSA change with age in men without cancer and compared this with men with screen-detected and clinically presenting cancer.³⁵ Limitations of this previous work include the use of only two cohorts of men with cancer and the derivation of the model on cancer-free men which meant that potential explanatory variables such as Gleason score, or age at diagnosis, were not included in the model. Initial examination of the comparison of observed PSA to that predicted for a cancer-free man of the same age indicated that this model might result in fewer alerts for clinical review.³⁵ However, clinical outcomes were not available for comparison. In order to study the effect of PSA change on clinical outcomes, it may be necessary to combine AM/AS cohorts, because very few deaths or cases of metastases have been reported at this stage.^{10,16,17}

Qualitative research in active monitoring/active surveillance

Qualitative methodology has not been widely applied in AM/AS research. Few studies have considered patients' and clinicians' views and experiences of AM/AS in depth, and the published literature largely refers to protocols that more closely follow AS than AM.

A considerable proportion of the qualitative literature relating to AM/AS considers men's treatment decision-making. Qualitative evidence largely places clinician recommendations and communication as principal drivers for patients' decisions to opt for or against AM/AS. This has tended to be framed in the literature as patients deferring the decision-making process to the clinician,^{39–42} or justifying their decision to opt for AM/AS in light of clinicians' recommendations.^{39,41,43} Similar decision-making tendencies are seen in earlier studies of localised PCa management that do not make specific reference to AM/AS.^{44,45} Of these studies, one suggests that passing on the decision-making power to the clinician can be an active choice on the man's part.⁴⁵ Qualitative interviews also suggest that men come to the decision to choose AM/AS by reading cues from their clinicians (e.g. descriptive terminology of the low-risk status of the cancer).^{39,41}

The qualitative findings provide insight into how clinicians may shape the decision-making process, but do not represent the prevalence of treatment decision-making preferences. A review of the literature found that men prefer to be actively involved in the treatment decision-making process.⁴⁶ Despite this, there is a substantial body of evidence that suggests that men's values and preferences are not reflected in the final treatment choices made.⁴⁷ A systematic review of treatment decision-making concluded that men's treatment choices tend to be reflective of information they receive, rather than their underlying values and preferences.⁴⁸

Only one qualitative study has considered AM/AS men's information needs, focusing broadly on psychosocial and educational needs. The single focus group study found that men had poor knowledge of their disease and tended to turn to the internet as their main source of information.⁴⁹ A number of studies have designed decision-support interventions to increase men's involvement in, and satisfaction with, decision-making in relation to treatment of PCa through increasing knowledge and education. Successful interventions have included providing audio-recordings of consultations, presenting individualised information to men and their partners,^{50,51} and providing counselling via nurses.⁵²

A handful of qualitative studies have sought to explore men's experiences of being on AM/AS. These have tended to focus on psychosocial issues, particularly issues of uncertainty. Two qualitative interview-based studies in Sweden and Canada report uncertainty as a major theme to emerge from interviews with men on AM/AS.^{53,54} Men talked about uncertainty across a number of domains, including disease progression, prognosis, symptom interpretation and morbidity associated with potential imminent treatment.

However, there was a suggestion that uncertainty was temporal, peaking at times of PSA testing or rebiopsy. Receipt of reassuring test results enabled men to 'continue with their normal lives'. The study by Oliffe *et al.*⁵³ also reported that men's strategies for coping with uncertainty included attempting to 'live a normal life' and to do all they can to assert control over their bodies (i.e. through lifestyle and diet changes, and use of complementary/alternative therapies).

Some qualitative studies have intentionally focused on the issue of uncertainty in AS, rather than identified this as a key theme through grounded theory approaches.⁵⁵ A study by Bailey *et al.* analysed men's accounts of being on AM/AS through the lens of a pre-existing theoretical model of uncertainty.⁵⁵ The authors identified uncertainty surrounding cancer progression and the questioning of whether or not one had made the 'right choice' as major themes to emerge from interviews. Linked with the uncertainty related to cancer progression was men's perceptions of how PSA monitoring was not a fail-safe indicator of disease status.⁵⁵

The issue of uncertainty has also been identified as a key issue through survey-based research. A survey asked men to provide open responses to a question that asked about the key advantages and disadvantages of being on AM/AS ($n = 103$).⁵⁶ One-third of men reported the risk of disease progression being a key disadvantage, one-quarter made non-specific reference to 'uncertainty' and 'distress' and 10% talked about the bothersome nature of regular follow-up.

Although research focusing on men's experience of AM/AS is limited, qualitative studies on decision-making have made some reference to men's AM/AS experiences. Davison *et al.* reported that men felt able to 'live a normal life', although the authors also emphasise patients' reported relaxed mentality.³⁹ A systematic review of psychological effects of AM/AS concluded that men could experience anxiety in response to the prospect of 'no intervention', uncertainty and lack of patient education/support. The authors suggest that these concerns are likely to be particularly relevant to men on AM/AS.⁵⁷ Overall, the literature on measures of anxiety/psychological distress and/or health-related quality of life suggests there are no significant negative outcomes associated with following an AM/AS programme,⁵⁸⁻⁶³ although this can depend on other factors (e.g. personality type, how men conceptualise the role of the clinician, marital status).^{59,64} However, it has also been reported that PSA testing and rebiopsies can raise anxiety levels in men on AM/AS, usually temporarily.⁶⁵ Furthermore, comparisons of psychological distress/health-related quality of life in men on AM/AS compared with men who undergo radical treatments is somewhat mixed.^{59,60,65-67}

To our knowledge, there has been no qualitative research that seeks to understand patients' and/or clinicians' experiences of AM/AS protocols. More generally, men's understanding of and views surrounding AM/AS protocols, including thresholds for moving off AM/AS, have not yet been reported. Furthermore, although guidance on AS protocols has been produced by authoritative bodies [e.g. the National Institute for Health and Care Excellence (NICE), the British Association of Urological Surgeons, the European Association of Urology], AM/AS protocols adopted by clinicians have not yet been explored in practice.

NHS policy and practice

The recently published 2014 NICE guidelines⁶⁸ suggest that men with low-risk localised PCa should be offered AS, which acknowledges that many men would not benefit from radical treatment. NICE recommends a protocol for the management of men on AS, including frequency of PSA testing, DRE and rebiopsy. NICE also suggests that AS could be offered to men with intermediate-risk localised PCa if these men do not wish to undergo immediate radical treatment. There is only a little information about treatments, including AS, as part of the NHS Prostate Cancer Risk Management Programme (see www.cancerscreening.nhs.uk/prostate/faq17.html).

Overview of this report

In *Chapter 2* we present a systematic review of AM/AS, which includes a summary of the included studies along with a meta-analysis into the factors affecting management change from AM/AS to radical treatment. A model for PSA change is developed in *Chapter 3*, using data from men following AM as part of the ProtecT trial. This model for PSA change is then validated in four external cohorts in *Chapter 4* and used to derive 95% reference ranges for PSA with age. *Chapter 5* compares the PSA reference range (PSARR) method against PSADT and PSAv in predicting several PCa outcomes in three external cohorts. *Chapters 6* and *7* summarise the qualitative interviews conducted with clinicians and patients about AM/AS and the active monitoring system (AMS). *Chapter 8* discusses several potential RCTs and summarises the issues surrounding any RCT of AM/AS strategies. *Chapter 9* contains recommendations for future researchers and health-care professionals in the area.

Chapter 2 Systematic review of active monitoring for prostate cancer

Aim

The aims of this stage of the project were to identify how AM/AS is carried out in ongoing studies worldwide and to assess whether or not any study protocols (such as eligibility) appeared to be beneficial for PCa outcomes. We also wished to investigate the PSA monitoring procedures used in modern AM studies, in order to compare these to PSARRs.

Search strategy

We conducted a systematic search of the online databases MEDLINE, EMBASE and Web of Science for articles/abstracts published between October 2004 (the end date of the previous systematic review by Martin *et al.*¹⁶) and April 2013. Reference lists of included articles and recent reviews were searched by hand for additional citations. A forward citation search of the five studies²⁰⁻²⁴ included by Martin *et al.*¹⁶ was also performed using the Web of Science database. Where potentially relevant conference abstracts only were available, authors were contacted to enquire about full publications of their work or to elicit further details of their AM/AS programme.

Inclusion criteria

Studies involving men with T1–T2 localised PCa, which was initially managed conservatively, were included if pre-defined clinical, pathological or biochemical criteria for clinical review were outlined. Studies involving recurrence after radical prostatectomy or radiotherapy (i.e. not initially managed conservatively) or non-clinical studies were excluded. We excluded any conference abstracts, non-systematic reviews, editorial comments, background papers or studies involving different treatments or diseases of the prostate. Where several papers reporting on the same cohort of men were found, the latest full analysis of the study was included. Title and abstract screening was undertaken independently by AJS, CM, KT and RMM. Each full paper was screened by AJS and independently by one of CM, KT and RMM.

Data extraction

Eligibility criteria, monitoring protocols, sample size, age, PSA, follow-up times, management change triggers, management change rates, metastases, PCSM and reasons for changing treatment were extracted manually from each paper by AJS and checked by one of CM, KT or RMM. Authors of publications found in our search were contacted to provide information where specific values were not given in a publication and to check that data extraction was correct. Management change rates are considered here as key short-term outcome measures for AS, whereas occurrences of metastases or PCa death are considered to be longer-term outcomes.

Data synthesis

Person-years per study were estimated as the median follow-up time multiplied by the sample size for each study; total person-years were the sum of these across included studies. A metaregression was conducted to investigate whether or not elements of study design were associated with the rate of management change per person-years. This rate was calculated for each study as:

$$\frac{\text{radical treatment events}}{\text{sample size} \times \text{median follow-up time}} \quad (1)$$

Heterogeneity between studies was measured using the I^2 statistic;⁶⁹ the higher the I^2 value, the more between-study heterogeneity is present. The covariates for the metaregression were (1) year of first recruitment; (2) whether or not a Gleason score of 3 + 3 or less was used for eligibility; (3) whether or not a PSA level of 10 ng/ml or less was used for eligibility; (4) the number of PSA tests in the first 3 years; (5) whether or not mandatory rebiopsy was specified in the protocol; and (6) whether or not PSA or PSA kinetic measures, such as PSADT or PSAv, were used to recommend clinical review or radical treatment. Together with univariable analysis, two multivariable metaregressions were performed grouping eligibility variables (1)–(3) and monitoring procedure variables (4)–(6). Average variable effects on management rate are presented alongside 95% confidence intervals (CIs).

Meta-analysis was performed in Stata 12⁷⁰ (StataCorp LP, College Station, TX, USA) to combine the effects of prognostic factors for management change. Factors were combined if they appeared in more than two studies that reported estimates of the association with change to radical treatment. Where applicable, a sensitivity analysis was used to check if the pooled combination of odds ratios and hazard ratios (HRs) led to different results from just pooled HRs or pooled odds ratios. We also used meta-analysis to study the reasons for management changes. Reasons for changing to radical treatment were given within studies and the proportions of each reason were combined to get an overall average estimate.

Bias in reporting

Publication bias could be an issue in our review, as we examined only abstracts published in English. However, we would expect that studies showing that AM/AS was not a safe option would be more likely to be published in English-language journals, and, therefore, if anything, this systematic review will overestimate the number of adverse events.

Results

Authors of potentially relevant conference abstracts were contacted to provide full papers if available, although we identified no publications by this route. We found 157 potentially eligible studies for full-paper screening, of which 38 papers met the inclusion criteria, reporting on 19 unique cohorts of men on AS (*Figure 1*).^{71–89} Two of these were updates of the five studies from the previous review by Martin *et al.*¹⁶ Updated cohort information was volunteered by one author.⁷⁸ We also included the three papers^{21–23} without update from the previous review, leading to 22 cohorts and 7111 men. The median age at entry ranged from 64 years to 70.5 years, with the youngest participant being 36 years⁸⁶ and the oldest being 97 years.⁷³ The median follow-up time between the cohorts was 3.7 years. Four studies had < 2 years' median follow-up^{23,77,88,89} and seven studies had > 5 years,^{21,71–73,79,80,85} up to a maximum of 7.5 years.⁷¹

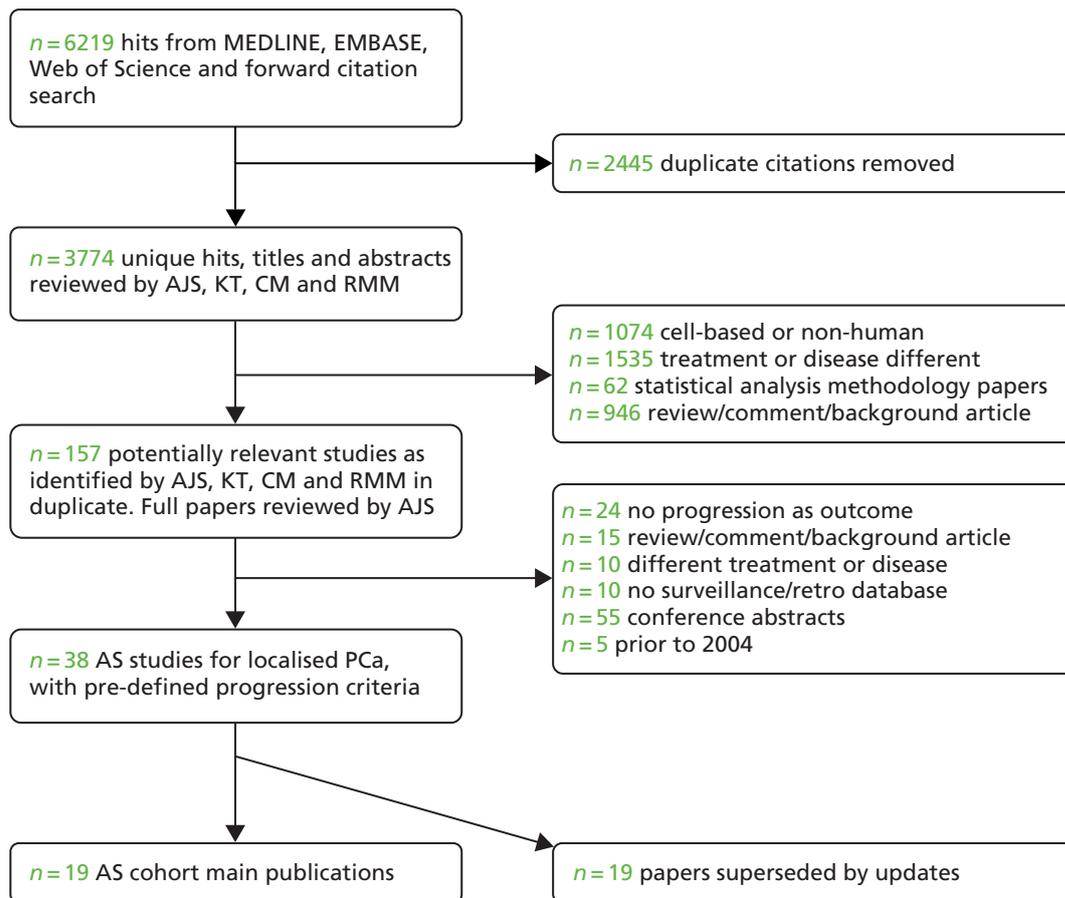


FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart of the screening process.

Eligibility criteria

Table 1 contains a summary of eligibility criteria, monitoring protocol and baseline characteristics. A combined Gleason score of 6 (3 + 3) or lower was required for entry into 12 of the 22 studies.^{71,72,74,77-79,82-84,87-89} A further five cohorts broadened this criterion to include a Gleason score of 7 (3 + 4),^{75,76,81,85,86} and two allowed up to a Gleason sum of 7,^{22,80} leaving three cohorts without an explicitly specified Gleason entry requirement.^{21,23,73} Additional pathology findings were used with some degree of consistency across the studies. Five studies allowed only those men with < 50% cancer in any single core into the study,^{74,77,82-84} with one study allowing only 20% involvement in any single core.⁷⁸ The number of positive cores allowed varied between an absolute number or a proportion; two studies required no more than half of the biopsy cores to be positive,^{85,88} whereas seven allowed two or fewer positive cores,^{74,78,82-84,86,89} and one study raised this to allow those men with three positive cores to enter.⁷⁷

Clinical tumour stage (T-stage) was used in 17 cohorts with thresholds of T2 (six cohorts^{75,76,80,81,83,89}), T2a (five cohorts^{74,77,85,87,88}), T1c (three cohorts^{23,82,84}), T1b⁷³ and T1a (two cohorts^{21,71}). There was no consensus among the 15 cohorts that used a PSA threshold for entry, with seven cohorts permitting men with PSA ≤ 10 ng/ml,^{74,77-79,86,87,89} three cohorts allowing PSA ≤ 15 ng/ml^{75,85,88} and three cohorts allowing PSA < 20 ng/ml.^{76,80,84} Two further studies used PSA density (PSAD) as an eligibility criterion, with limits of PSAD < 0.15 ng/ml/cc⁸² and PSAD < 0.2 ng/ml/cc.⁸⁹ Seven studies did not explicitly use PSA for eligibility,^{21-23,71-73,83} although six of these involved men diagnosed between 1983 and 1991^{21-23,71-73} (i.e. in the pre- and early-PSA-screening era). Age restrictions were used in just four studies.^{74,79,84,88}

TABLE 1 Eligibility criteria, monitoring protocol and characteristics for the 22 included studies, sorted by year of commencing recruitment

Setting	Study design				Monitoring protocol			Patient baseline characteristics		
	Eligibility				PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)
	PSA	Gleason score ≤	Tumour stage ≤	Other						
Taichung Veterans General Hospital, Taichung City, Taiwan ²¹	-	-	1a	TURP for benign hyperplasia	Every 3-6 months (after 1990)	Every 3-6 months	As a result of DRE/PSA	52 (1983-96)	Mean 71.5 (SD 7)	Not specified
Memorial Sloan Kettering Cancer Center, New York, NY, USA ²²	-	4+3	-	No significant comorbidities; eligible for radical treatment	Every 3 months for 1 year, then every 6 months	Every 3 month for 1 year, then every 6 month	At 6 months or as a result of DRE/PSA	88 (1984-2001)	Mean 65.3 (range 44-79)	5.9 (range 0.09-30.2)
Hospital Universitario Miguel Servet, Zaragoza, Spain ⁷¹	-	3+3	1a	-	Every 6 months	Every 6 months	Restaging biopsy post diagnosis then as a result of DRE/PSA	16 (1986-99)	Mean 64 (range 49-71)	Not specified
McGill University, Montreal, QC, Canada ⁷²	-	3+3	-	Patient choice, limited life expectancy, presumed insignificant cancer	Every 3-6 months	Every 3-6 months	Every 12 months or as a result of DRE/PSA	186 (1987-2006)	67 (range 49-78)	<4 (23%), 4-10 (65%), 10-20 (10%), >20 (2%)
Cochin Hospital, Paris, France ⁷³	-	-	1a	-	Within 2 months then every 6 months for 2 years, then every 12 months	-	As a result of PSA	144 (1988-2005)	Mean 69 (range 52-97, SD 9.2)	7 (range 0.6-29)

Study design		Monitoring protocol			Patient baseline characteristics				
Eligibility		PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)		
Setting	Gleason score \leq	Tumour stage \leq	Other	PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)
University of Connecticut Health Center, Farmington, CT, USA ⁷⁴	3+3	2a	≤ 2 positive cores; $< 50\%$ cancer in any core; age < 75	Every 6 months or every 3 months if PSA increasing	Every 6–12 months	Within 2 years or as a result of DRE/PSA or patient choice	40 (1990–2006)	68 (range 52–75)	6.3 (range 4–9.1)
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA ²³	–	1c	–	At 3 months then every 6 months	At 3 months then every 6 months	Not routinely taken	27 (1991–6)	Mean 65 (SD 5)	Mean 5.3 (range 2–24, SD 9.8)
ERSPC, Rotterdam, the Netherlands ⁷⁵	4+3	2	–	Chart reviews every 6 months	Chart reviews every 6 months	Not routinely taken	278 (1993–9)	69.8 (IQR 66.1–72.8)	3.6 (IQR 3.1–4.8)
Royal Marsden NHS Foundation Trust, London, UK ⁷⁶	4+3	2	Fitness for radical treatment	3–6 months, then every 6 months after 2 years	3–6 months, then every 6 months after 2 years	Not routinely taken	80 (1993–2002)	70.5 (range 59–81)	< 4 (20%), 4–10 (52%), 10–20 (25%), > 20 (1%)
Memorial Sloan Kettering Cancer Center, New York, NY, USA ⁷⁷	3+3	2a	≤ 3 positive cores (out of at least 10); $< 50\%$ cancer in any core; confirmatory biopsy before starting surveillance	Every 6 months	Every 6 months	Within 12–18 months, then every 2–3 years or as a result of DRE/PSA	238 (1993–2009)	64 (IQR 58–68)	4.1 (IQR 2.5–5.6)
University of Miami, Coral Gables, FL, USA ⁷⁸	3+3	2a	≤ 2 cores; $\leq 20\%$ cancer in any core	Every 3–4 months for 2 years, then every 6 months	Every 3–4 months for 2 years, then every 6 months	Within 1 year then every 1–2 years	276 (1994–2011)	Mean 63 (range 42–79)	4.8

continued

TABLE 1 Eligibility criteria, monitoring protocol and characteristics for the 22 included studies, sorted by year of commencing recruitment (continued)

Setting	Study design			Monitoring protocol			Patient baseline characteristics		
	PSA	Gleason score ≤	Tumour stage ≤	PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)
University of Toronto, Toronto, ON, Canada ⁷⁹	≤ 10, 2000–1995–1999	3 + 3 2000–1995–1999	–	Every 3 months for 2 years, then every 6 months for stable patients	–	Within 6–12 months then every 3–4 years or as a result of PSADT	450 (1995–2010)	< 70 (48.7%), ≥ 70 (51.3%)	0–2.5 (12%), 2.5–5 (25%), 5–10 (48%), 10–15 (12%), 15–20 (2%), n/a (0.4%)
ERSPC, Gothenburg, Sweden ⁸⁰	< 10 (low-risk group)	3 + 3 (low-risk group)	1 (low-risk group)	Every 3–6 months	Every 3–6 months	As a result of signs of PSA or T-stage progression	439 (1995–2010)	65.4 (51–70)	Not specified
University of California, San Francisco, San Francisco, CA, USA ⁸¹	Within CAPRA score	3 + 3 (low-risk group) 3 + 4 (intermediate-risk group)	2	Every 3 months	Every 3 months	Every 12–24 months	466 (1995–2010)	Mean 62.8 (SD 8.1)	Low risk 4.99 (range 0.3–11); intermediate risk 10.3 (range 3.14–37.91)
JH University, Baltimore, MD, USA ⁸²	PSAD < 0.15 ng/ml/cc	3 + 3	1c	Every 6 months	Every 6 months	Every 12 months	769 (1995–2011)	66 (range 45–92)	5 (range 0.3–19) in men eventually treated; median 4.7 (range 0.2–24) in untreated

Study design			Monitoring protocol				Patient baseline characteristics		
Eligibility			PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)	
Setting	Gleason score \leq	Tumour stage \leq	Other	PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)
Harvard University, Cambridge, MA, USA ⁸³	3+3	2c	≤ 2 positive cores; $< 50\%$ cancer in any core	Every 6 months	Every 6 months	Every 12–18 months	135 (2000–10)	Not specified	Not specified
Multi-institutional, Kagawa, Japan-based ⁸⁴	3+3	1c	≤ 2 positive cores $< 50\%$ cancer in any core Age, 50–80 years	Every 2 months for 6 months, then every 3 months	–	At 12 months	118 (2002–3)	50–59 (5%), 60–69 (43%), 70–74 (37%), 75–80 (16%)	Mean 7.2, 81% ≥ 10
Royal Marsden NHS Foundation Trust, London, UK ⁸⁵	3+4	2	$\leq 50\%$ positive cores Age, 50–80 years	Every 3 months in first year, every 4 months in second year, every 6 months after 2 years	Every 3 months in first year, every 4 months in second year, every 6 months after 2 years	Within 18–24 months then every 2 years	471 (2002–11)	66 (range 51–79)	6.4 (range 0.2–14.5)
Southern Health, Melbourne, Australia ⁸⁶	3+4	–	≤ 2 positive cores; < 3 mm in each core of cancer T1a if TURP diagnosed	Every 3 months	Every 6 months	Within 12–18 months then every 3 years	154 (2003–10)	63 (range 36–81)	Mean 6.5 (range 0.3–22)
Cleveland Clinic, USA ⁸⁷	3+3	2a	Confirmatory rebiopsy within 6 months	Every 6–12 months	Every 6–12 months	At least every 2 years	89 (2004–9)	67 (range 52–83)	4.8 (range 0.6–18.4)

continued

TABLE 1 Eligibility criteria, monitoring protocol and characteristics for the 22 included studies, sorted by year of commencing recruitment (continued)

Setting	Study design			Monitoring protocol			Patient baseline characteristics			
	Eligibility			Monitoring protocol			Patient baseline characteristics			
	PSA	Gleason score \leq	Tumour stage \leq	Other	PSA	DRE	Biopsy	Sample size (years recruited)	Median age (years)	Median PSA (ng/ml)
Frimley Park Hospital, Frimley, UK ⁸⁸	≤ 15	3 + 3	2a	$\leq 50\%$ positive cores; ≤ 10 mm of disease in single core; age ≤ 75 years	Every 3 months	–	Within 12 months then every 18 months	101 (2006–10)	68 (range 51–75)	6.4 (range 0.9–15)
PRIAS (International), Rotterdam-based, Holland ⁸⁹	≤ 10 ; PSAD < 0.2 ng/ml/cc	3 + 3	2	≤ 2 positive cores	Every 3 months for 2 years, then every 6 months	Every 6 months for 2 years, then every 12 months	1, 4 and 7 years or annually if $3 \leq$ PSADT ≤ 10	2494 (2006–12)	65.8 (IQR 61.0–70.4)	5.6 (IQR 4.4–7.0)

CAPRA, CAncer of the Prostate Risk Assessment; ERSPC, European Randomized Study of Screening for Prostate Cancer; IQR, interquartile range; PRIAS, Prostate Cancer Research International: Active Surveillance; SD, standard deviation; TURP, transurethral resection of the prostate.

Surveillance or monitoring protocol

Serial PSA measurement as a measure for surveillance or monitoring during follow-up was common to all studies, but the frequency of measurement was not consistent. Eleven studies maintained PSA testing every 3–6 months throughout;^{21,71,72,75,77,80–83,86,88} eight reduced the frequency of PSA testing after men had been on surveillance beyond 1 or 2 years.^{22,23,73,76,78,84,85,89} Frequent testing occurred in the Kagawa and modern Royal Marsden Hospital (RMH) studies, with men scheduled for 22⁸⁴ and 18⁸⁵ PSA tests, respectively, in the first 3 years.

Digital rectal examination was regularly performed in all but five^{73,79,84,86,88} of the studies, with inconsistency of frequency between the study protocols. In the 17 studies with both DRE and PSA testing, 14 assessed both at the same frequency, whereas the other three had DRE less often than PSA tests.^{74,85,89}

Repeat biopsy testing was used by 19 studies, with 16 having a scheduled biopsy within 2 years of beginning surveillance.^{22,74,77–79,81–89} Biopsies were repeated at least every 2 years in nine studies.^{72,78,81–85,87,88} Three studies undertook repeat biopsy only where worsening PSA or DRE results were evident or at the request of the patient or clinician.^{21,73,80} Three studies did not include routine rebiopsy or rebiopsy as a result of worsening DRE or PSA.^{23,75,76}

Triggers for clinical review or radical treatment

Triggers for recommending clinical review and radical treatment are summarised in *Table 2*. Almost all studies used histological findings such as an increase in Gleason score (18 cohorts^{22,71–74,77–89}), number of positive cores or percentage of cancer on a single core (15 cohorts^{22,71–74,77–79,82–86,88,89}) to recommend radical treatment. Less commonly, T-stage (six cohorts^{77,79,80,82,84,89}) and PSA (13 cohorts^{21–23,76,77,79–82,84–86,89}) were used as triggers to recommend rebiopsy or radical treatment. In four studies, a threshold of PSADT was used, with this being 2 years in two of those studies^{81,84} and 3 years in the other two studies.^{79,86} In one study, radical treatment was offered to men with a PSA velocity > 1 ng/ml/year,⁸⁵ while in another such men were rebiopsied.⁷¹ In two studies, a combination of PSA kinetics with other factors such as DRE and rebiopsy findings was used to recommend radical intervention.^{22,87}

Radical treatment rates

In 17 studies, up to 33% of men received radical treatment after a median of 3.6 years.^{21,23,71–76,78,79,82,83,85–89} In one of these studies, all 16 men remained on monitoring after 7.5 years, having started with very low-risk T1a disease.⁷¹ The University of California, San Francisco intermediate-risk group radically treated 35% of men after 3.9 years;⁸¹ the early Memorial Sloan Kettering study radically treated 35% of men after 3.7 years;²² and the later Memorial Sloan Kettering cohort changed the treatment of 36% of men.⁷⁷ In Gothenburg, 37% of men received radical treatment after a median follow-up of 6 years.⁸⁰ The highest proportion of men who changed from AS to radical treatment was in Kagawa, with 47% leaving surveillance for radical treatment by 4.5 years of follow-up.⁸⁴

Study factors associated with the rate of management change

Total person-years were estimated to be 22,545 in the 22 cohorts. The average rate of change of management (found through metaregression) was 84 (95% CI 61 to 106) per 1000 person-years. In the context of the follow-up found here (i.e. roughly 4 years), we estimate that 84 out of 250 men on monitoring would receive radical treatment during four years of follow-up (*Figure 2*). There was evidence ($R^2 = 96\%$; $p < 0.0005$) that the rate of change differed substantially between cohorts. The minimum rate of change was 11 per 1000 person-years²¹ and the maximum was 218 per 1000 person-years.⁸⁸

The rate of management change was greater in studies with recruitment periods that started in more recent years; for each calendar year increase in start date, an extra four men changed treatment per 1000 person-years (95% CI 1 to 7 treatment changes per 1000 person-years; $p = 0.014$). Admitting men with higher Gleason scores was associated with an average decrease of 50 treatment changes per 1000 person-years (95% CI 9 to 88 treatment changes per 1000 person-years; $p = 0.025$), compared with studies with eligibility criteria including Gleason score of ≤ 6 . Studies that had a mandatory scheduled

TABLE 2 Management change definitions and results for the 22 included studies

Setting	Trigger for clinical review and management change criteria	Median follow-up years (range)	Percentage changing to radical treatment (median time, years)	Percentage 'progressed' objectively (median time, years)	Progression-free probability	Treatment-free probability	Reasons for changing treatment
Taichung Veterans General Hospital, Taichung City, Taiwan ²¹	Abnormal DRE or progressive elevation of PSA	7.3 (0.5–15)	8 (2.5)	8	–	–	All four had abnormal DREs and elevation of PSA
Memorial Sloan Kettering Cancer Center, New York, NY, USA ²²	Three or more points coming from system including Gleason score increase, PSAv > 0.75 ng/ml/year, DRE/TRUS findings, biopsy findings	3.7 (0.6–13.5)	35 (7.3)	25 (3.75)	95% at 1 year, 88% at 2 years, 67% at 5 years, 56% at 10 years	93% at 1 year, 84% at 2 years, 58% at 5 years, 41% at 10 years	17 had three or more points Seven scored two and were anxious
Hospital Universitario Miguel Servet, Zaragoza, Spain ⁷¹	PSA > 4 or PSAv > 1 ng/ml/year leads to rebiopsy where upgrade can lead to treatment	7.5	0	0	–	–	Seven patient choice Not specified
McGill University, Montreal, QC, Canada ⁷²	Any of: predominant Gleason score of 4 pattern; > 2 positive cores; > 50% cancer/core, ≥ T2b	6.3 (1.7–14.1)	16 (3.7)	18	82% at 5 years for negative first biopsy group, 50% at 5 years positive first biopsy	–	Nine patients ≥ T2b 13 > 50% cancer/core Five Gleason score 22 > 3 positive cores
Cochin Hospital, Paris, France ⁷³	Doubling of PSA from baseline leads to rebiopsy where upgrade can lead to treatment	5.1	19	21 (mean 5.1)	75% at 5 year	–	Two patient choice Not specified

Setting	Trigger for clinical review and management change criteria	Median follow-up years (range)	Percentage changing to radical treatment (median time, years)	Percentage 'progressed' objectively (median time, years)	Progression-free probability	Treatment-free probability	Reasons for changing treatment
University of Connecticut Health Center, Farmington, CT, USA ⁷⁴	Any of: increase in Gleason score; increase in number of positive cores; onset of urinary symptoms; change in DRE; patient request	4 (1–14)	23 (2.75)	Not specified	–	74% at 5 year	Not specified
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA ²³	Abnormal DRE or three consecutive PSA increases with total increase of 5 ng/ml	Mean 1.9 (0.5–5.1)	19 (1)	33	80% at 1 year	–	Five PSA increases
ERSPC, Rotterdam, the Netherlands ⁷⁵	Patient desire and/or clinician advice	3.4	29 (2.5)	Not specified	–	86% at 2 years, 78% at 4 years, 66% at 6 years, 53% at 8 years, 31% at 10 years	Not specified
Royal Marsden NHS Foundation Trust, London, UK ⁷⁶	Any of rate of PSA increase, subjective decision by patient and clinician	3.5 (0.1–9.7)	14	11 (1.1)	–	79.2 at 5 years (95% CI 63.9 to 88.6)	Nine based on rate of PSA rise Two patient choice
Memorial Sloan Kettering Cancer Center, New York, NY, USA ⁷⁷	No longer meet study criteria (i.e. any of PSA \geq 10, Gleason score of \geq 7, > 3 positive cores, > 50% tumour/core, > T2a)	1.8 for untreated; 11% \geq 5 years	36	26	80% at 2 years, 60% at 5 years	–	34 PSA \geq 10 23 Gleason score of \geq 7 Seven positive cores > 3 One T-stage Two tumour/core > 50% (Five men had two or three of these factors) 25 patient choice

continued

TABLE 2 Management change definitions and results for the 22 included studies (continued)

Setting	Trigger for clinical review and management change criteria	Median follow-up years (range)	Percentage changing to radical treatment (median time, years)	Percentage 'progressed' objectively (median time, years)	Progression-free probability	Treatment-free probability	Reasons for changing treatment
University of Miami, Coral Gables, FL, USA ⁷⁸	Any of Gleason score of ≥ 7 , increase in number of positive cores, increase in percentage of cancer/core or personal choice	3.3 (1–17.3)	26	22	–	85.7% at 5 years	15 Gleason 24 tumour volume (i.e. no. cores or percentage cancer/core) 28 both tumour volume and Gleason Eight patient choice
University of Toronto, Toronto, ON, Canada ⁷⁹	Any of PSADT < 3 years, histological upgrade on rebiopsy or clinical progression	6.8	30	30	–	84% at 2 years, 72% at 5 years, 62% at 10 years	65 patients PSADT < 3 years 36 Gleason increase Six T-stage Four volume Two urethral obstruction 14 patient choice Eight unknown

Setting	Trigger for clinical review and management change criteria	Median follow-up years (range)	Percentage changing to radical treatment (median time, years)	Percentage 'progressed' objectively (median time, years)	Progression-free probability	Treatment-free probability	Reasons for changing treatment
ERSPC, Gothenburg, Sweden ⁸⁰	Established PSA, stage or Gleason score progression	6 (0.08–15.1)	37	Not specified	–	76% at 2 years, 61.5% at 5 years, 45.4% at 10 years, 37.1% at 14 years	77 Gleason Seven T-stage 45 PSA
University of California, San Francisco, San Francisco, CA, USA ⁸¹	Upgrade to PSADT ≤ 2 years, Gleason score of $\geq 4 + 3$ if already 3 + 4 or to 3 + 4 otherwise	3.9 (1–15.2)	30% of low risk (4) 35% of intermediate risk (4)	39	54% at 4 years for low risk 61% at 4 years for intermediate risk	–	Four anxiety Four other symptoms 73 reclassified on repeat biopsy 17 on PSADT 38 on PSA
JH University, Baltimore, MD, USA ⁸²	Any of: PSAD ≥ 0.15 ml/cc, Gleason score of ≥ 7 , > 2 positive cores, > 50% tumour/core	2.7 (0.01–15)	33 (2.2)	24	–	81% at 2 years, 59% at 5 years, 41% at 10 years	Nine other pathological 11 other 91 number of positive cores or tumour involvement 106 Gleason 67 patient choice

continued

TABLE 2 Management change definitions and results for the 22 included studies (continued)

Setting	Trigger for clinical review and management change criteria	Median follow-up years (range)	Percentage changing to radical treatment (median time, years)	Percentage 'progressed' objectively (median time, years)	Progression-free probability	Treatment-free probability	Reasons for changing treatment
Harvard University, Cambridge, MA, USA ⁸³	Any of: Gleason score of ≥ 7 , > 3 positive cores, > 50% tumour/core	2.4	27	27	95% at 1 year, 82% at 2 years, 72% at 3 years, 59% at 4 years, 45% at 5 years	-	18 number of cores 11 Gleason Six number of cores + Gleason score 17 PSADT ≤ 2 years
Multi-institutional, Kagawa, Japan based ⁸⁴	PSADT ≤ 2 years or pathological upgrade on rebiopsy	4.5	47	29	-	49% at 3 years	One T-stage change 16 pathology change 15 patient choice Eight comorbidities Seven unknown
Royal Marsden NHS Foundation Trust, London, UK ⁸⁵	Any of PSAv > 1 ng/ml/year, Gleason score of $\geq 4 + 3$, > 50% tumour/core	5.7	31 (5.4)	22	-	89% at 2 years (95% CI 86% to 92%), 70% at 5 years (95% CI, 65% to 75%)	18 histological alone 56 PSAv alone 23 both histological and PSAv 40 patient choice

Setting	Trigger for clinical review and management change criteria	Median follow-up years (range)	Percentage changing to radical treatment (median time, years)	Percentage 'progressed' objectively (median time, years)	Progression-free probability	Treatment-free probability	Reasons for changing treatment
Southern Health, Melbourne, Australia ⁸⁶	PSADT < 3 leads to rebiopsy, any Gleason score or mm cancer/core increase	2.4 (0.2–7.9)	16 (2.4)	17	–	62% at 5 years, 45% at 10 years	26 reclassified on repeat biopsy
Cleveland Clinic, USA ⁸⁷	Either Gleason score of ≥ 7 or a combination of PSA, PSA kinetics, DRE cancer quantity, Gleason score; never PSA alone	2.75 (IQR 1.7–3.75)	18	39	93% at 3 years (95% CI 85% to 97%)	87% at 3 years (95% CI 78% to 93%)	One patient choice 10 patient choice Six increased Gleason score
Frimley Park Hospital, Frimley, UK ⁸⁸	Any of: upgrade on TTB increase in Gleason score, > 50% positive cores, ≥ 10 mm tumour in 1 core	1.5 (1–2.3)	33	33	–	–	One percentage of cancer/core One number of cores 25/34 reclassified on TTB chose treatment
PRIAS (International), Rotterdam based, the Netherlands ⁸⁹	Any of PSADT < 3 years, Gleason score of ≥ 7 , > 2 positive cores, stage > T2	1.6 (IQR 1.0–2.8)	21	16	–	77% at 2 years, 68% at 4 years	10 patient choice 387 patients on biopsy or PSADT 47 anxiety, 93 other

ERSPC, European Randomized Study of Screening for Prostate Cancer; IQR, interquartile range; PRIAS, Prostate Cancer Research International: Active Surveillance; TRUS, transrectal ultrasound scan; TTB, transperineal template biopsy.

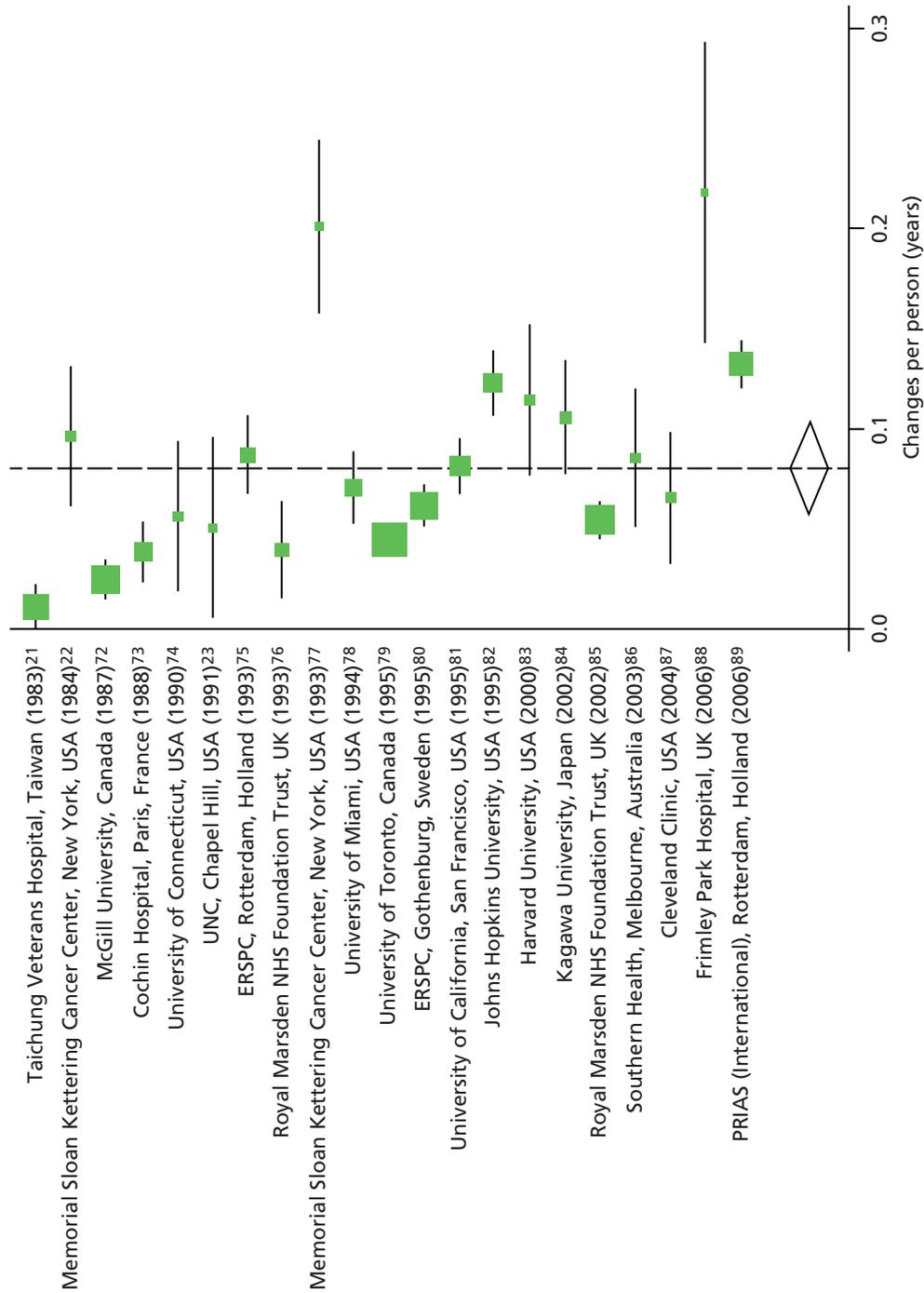


FIGURE 2 Forest plot of the proportion of men changing to radical treatment per person-years, sorted by year of beginning recruitment. ERSPC, European Randomized Study of Screening for Prostate Cancer; PRIAS, Prostate Cancer Research International; Active Surveillance; UNC, University of North Carolina.

rebiopsy had, on average, 50 more men change treatment per 1000 person-years (95% CI 7 to 94 treatment changes per 1000 person-years; $p = 0.025$) compared with those that rebiopsied on worsening results from PSA tests or DRE only. Each of these associations was found in univariable analyses. Two multivariable meta-regressions were performed to account for mutual confounding. In these models, the results were slightly attenuated, although there still remains some evidence for associations with Gleason score of $> 3 + 3$ (46 fewer changes, 95% CI -5 to 96 fewer changes; $p = 0.07$) and regular biopsy (49 more changes, 95% CI -5 to 98 more changes; $p = 0.07$).

Reasons for changing management

Eighteen studies reported reasons men had for changing treatment.^{21–23,72,76–89} Up to an average of 37% of changes to radical treatments were a result of a reclassified Gleason score (95% CI 23% to 51%),^{72,77–89} whereas an average of 38% were caused by men meeting PSA, PSADT or PSAv triggers (95% CI 25% to 51%).^{76,77,79–81,84,85,89} Other tumour-related reclassifications, such as the number of positive cores, caused an average of 23% of the management changes (95% CI 8% to 37%).^{72,77–79,81–83,89} Among those studies detailing reasons, an average of 21% (95% CI 13% to 30%) of treatment changes were a result of patient choice or anxiety.^{22,72,76–80,82,84–89} Not all of these four types of reasons (Gleason score, PSA, other tumour and patient choice) were possible for all studies and, hence, the sum of the given averages is over 100%. For example, if study 1 has 100 men who change management from AS to radical treatment, with 50 changing as a result of Gleason score reclassification and 50 changing as a result of a rising PSA, then each reason contributes 50%. In study 2, 100 men receive radical treatment; however, in this study, men can only receive radical treatment as a result of Gleason score. Thus, averaging the two studies, we can either say that Gleason score contributes 75% [= (50% + 100%/2)] and PSA 25% [= (50% + 0%)/2] of the reason why men receive radical treatment or we can ignore the PSA contribution from study 2 (i.e. 0%) and calculate that Gleason score contributes 75% [= (50% + 100%/2)] and PSA contributes 50% (= 50%/1) of the reasons for a change in management. We present the second method of reporting, which results in a total value over 100%.

Prostate-cancer-specific mortality and metastases

Overall, eight PCa deaths were reported among the 7111 men and 22,545 person-years of follow-up. In Gothenburg, one man changed to hormone therapy after 8.6 years on surveillance; he developed distant metastases and died 12.7 years after diagnosis.⁸⁰ Five PCa-specific deaths occurred in the Toronto cohort, with all having a PSADT < 2 years, such that triggers for intervention were met and treatment recommended; two men received radiation therapy, one underwent radical prostatectomy and two refused treatment. The deaths occurred 3.7, 5.2, 5.3, 8.7 and 9.6 years after diagnosis.⁷⁹ In the RMH cohort, two men died from PCa 4 years and 8 years after diagnosis, respectively.⁸⁵ Both had a PSAv of > 1 ng/ml/year, subsequent Gleason score upgrade on rebiopsy and received androgen-deprivation therapy. Six further cases of metastases have been reported among the remaining men.^{21,72,80,89}

Summary

We reviewed 22 studies of men with localised PCa on AS, more than four times the number published in our previous review.¹⁶ Follow-up times were generally short, with 15 studies having < 5 years median follow-up. Eight PCa deaths and five further cases of metastases were found in 7111 men in short-term follow-up.

Eligibility and triggers for management change vary across studies, leading to different study populations and difficulty in making comparisons between studies. All studies used regular PSA testing, combined with repeat biopsy testing in 19 of the 22 studies. An increase in Gleason score was the most common trigger for recommendation of switching to radical treatment.^{19,23–27,30–41} The frequencies of PSA testing, DRE and rebiopsy during surveillance were inconsistent across the studies. The rate of change from AM/AS to radical treatment was found to be 30% or less in 15 of the 22 studies. The primary cause of management change was an upgraded Gleason score. In addition, across studies describing reasons for management

change, an average of 21% of men who changed from AS to radical treatment did so without meeting triggers for clinical review.

Studies including men with a Gleason score of $> 3 + 3$ had 50 fewer men changing per 1000 person-years (95% CI 9 to 88 fewer men changing per 1000 person-years). One explanation could be that those with a Gleason score of 6 have an increased chance of receiving an upwards reclassification compared with those with an original Gleason score of 7. Many men who receive a Gleason score of 6 at diagnosis will be undergraded,⁹⁰⁻⁹² such that on rebiopsy they are deemed to have progressed. Another reason could be that those studies with Gleason score of $3 + 4$ eligibility have higher treatment-switching thresholds.

Studies with scheduled rebiopsy were shown to have an increased rate of management change. This is to be expected given that rebiopsy offers the opportunity to detect small areas of high-grade tumour missed by initial biopsies, particularly among men with the most common $3 + 3$ Gleason score baseline results. It is well established from surgical series that such upgrading (as opposed to true progression) at a rate of around 25–30% is to be expected.⁹⁰⁻⁹²

A higher baseline PSA and T-stage were also found to be associated with a higher rate of change to radical treatment. PSA change in its various forms (doubling time, velocity, etc.) was also responsible for many changes of management, although evidence for its veracity in this regard is lacking.^{25,28}

Further follow-up data, which will include definitive outcomes, are needed to clarify the long-term impact of AM/AS, in terms of management change and PCa mortality. The use of the word 'progression' by several studies to mean a worsening condition rather than treatment switching is misleading. Furthermore, not all studies explicitly provided a breakdown of what caused men to leave surveillance and switch to radical treatment, making comparisons difficult. In the metaregression, explicit information about those leaving treatment was not available for all studies. Thus, all those receiving radical treatment were grouped together and not separated by the reason for changing treatment, the radical treatment received or PCa outcome. To this end, it would be beneficial in future publications of ongoing AM/AS cohorts for detailed reasons for management change to be given. A breakdown of radical treatment received by reason for leaving monitoring would also be helpful in determining whether or not men opt for radical treatment or WW, particularly as they age.

Chapter 3 Building a model for prostate-specific antigen during active monitoring

The material from this chapter has been published as an article.⁹³

Aims and background

This part of the project involved:

- the development of a model for PSA change with age in one cohort of men with clinically localised PCa
- deciding whether or not baseline diagnostic variables or characteristics improve the model.

Here, we present the decisions made in formulating a model for age-related PSA change in men with PCa. We introduce the AM data set and detail the methods to be compared (see *Methods*). In *Results*, we summarise the modelling of PSA by each approach and in *Choosing a model for PSA reference ranges* we choose the optimum method for PSARRs.

Methods

Study population

The ProtecT study is a large UK-based multicentre clinical trial comparing surgery, radiotherapy and AM in the treatment of screen-detected (PSA ≥ 3 ng/ml) and clinically localised PCa.^{14,15} Data were available from 512 men who declined random allocation of their treatment but chose to be managed according to the AM protocol. They had been followed over an average period of 4.8 years [standard deviation (SD) 2.4 years] with a median of 14 visits per individual (7438 total PSA tests). Covariates collected include Gleason score (or grade),⁹⁴ a histological grading system for PCa, which allows for the two main patterns of tumour-cell structure evident upon biopsy to be graded and combined into a single score. All participants in this cohort have a Gleason score of 6 or 7; however, a Gleason score of 7 can refer to either a 3 + 4 (pattern 3 predominant 4, also present) or a 4 + 3.

We have centred age at 50 years (i.e. considering age 50 years to be the baseline age), given that the inclusion criterion for ProtecT is men aged between 50 and 69 years old. Follow-up of men is irregular, as both the duration and timing of blood tests vary across men and, because of this, classical longitudinal approaches for balanced designs are inappropriate. Furthermore, age, rather than time in the study, is used as the covariate over which change in PSA is estimated. Using age makes the model more portable to populations where age at, and method of, diagnosis for PCa may vary. Furthermore, as there is a great deal of variation in when a diagnosis is made, and, consequently, in when treatment is started in relation to the natural history of the disease, and as no PSA altering treatment is used in this cohort of men, changes in PSA level will be aligned with age rather than the date of diagnosis. Here, given that we are interested in model comparison, we treat Gleason score as a binary variable (score of 6 vs. 7) for simplicity.

Prostate-specific antigen has a positively skewed distribution and, for this reason, a log transformation of the response is the usual initial step in its analysis.^{29,35,95,96} Figure 3 displays simulated PSA trajectories for five men on AM, the high variation in PSA level and steady upwards trend being typical of a real data series from men on ProtecT. These data are simulated, because presentation of post-randomisation results from the trial is restricted until primary analysis in 2015.

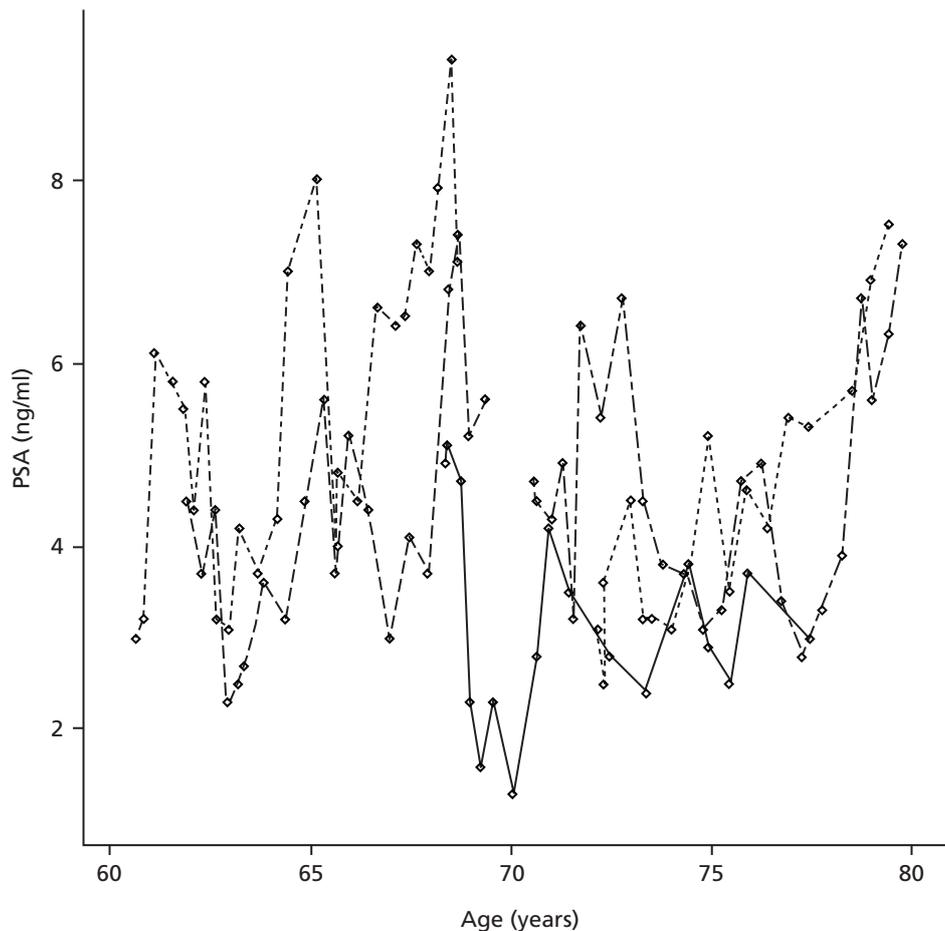


FIGURE 3 Plot of PSA trajectories for five simulated men.

Methods for modelling repeated measures of prostate-specific antigen

Four approaches were used to fit the PSA data from ProtecT, namely linear mixed models (LMMs), fractional polynomial mixed models (FPMs), regression spline mixed models (RSMMs) and principal components analysis through conditional expectation (PACE). These methods are described in detail in Appendix 4.

Prostate-specific antigen reference ranges

A suitable model for PSARRs must satisfy two criteria. First, it must offer an accurate description of the trend of PSA on average and in individuals. Second, it must be able to make accurate predictions about new PSA observations for an individual and about the entire PSA trajectory for a new individual. Once the model is chosen, 95% of observations will be below:

$$\widehat{\text{PSA}} + 1.96\sqrt{\text{var}(\text{PSA})}. \quad (2)$$

We are only interested in an upper bound for PSA, given that the method would be used to alert men with high observations of PSA.

Conditional predictions for multilevel models

Using methods developed for LMMs by Pan and Goldstein⁹⁷ and Tilling *et al.*,⁹⁸ we can enhance predictions for future observations by conditioning on already observed PSA. This allows us to make better use of models when moving to a different population of interest or for new individuals from the same population. To predict a future PSA for an individual, conditional on a previously observed PSA, we regress the deviation from the fixed part of the model at the second measurement on the observed deviation from the model at the previous measurement. Mathematically, to condition the prediction of PSA₂ on PSA₁:

$$\widehat{PSA}_2^{(c)} = \widehat{PSA}_2 + \frac{\text{cov}(PSA_1, PSA_2)}{\text{var}(PSA_1)} \times (PSA_1 - \widehat{PSA}_1). \quad (3)$$

Results

Linear mixed models, FPMs and RSMMs were estimated in Stata⁷⁰ and the PACE method was fitted using MATLAB version 7.14 (R2012a) (The MathWorks Inc., Natick, MA, USA).⁹⁹ The resulting models were used to obtain fitted values of PSA using both unconditional and conditional methods where applicable. Prediction performance was measured using the root-mean-square error (RMSE), whereas model fit was assessed using Akaike information criterion (AIC) and the residual error σ_ϵ^2 .

Linear mixed model

A linear mixed model was used to estimate the linear effect on log(PSA) of age, Gleason score and their interaction. The fixed parameter estimates are given in *Table 3*. Those with a higher Gleason score have an almost 30% increased exponential growth rate per year in their PSA (i.e. from 0.073 to 0.106). Given that the model must follow a log-linear trend, the mean LMM fit is lower to begin with in order to fit the data across later ages where PSA is generally higher.

The interpretation of this model is straightforward for use in a clinical setting. Its simplicity also allows for easy transfer to different populations, where more complex models may not fit well as a result of, say, knot placement or polynomial degree being too sample specific. This model could be used to predict PSA trajectory for a newly diagnosed man aged between 50 and 70 years at diagnosis, given his Gleason score. Once an initial PSA is observed, the model can be updated by conditioning on this new observation, thus making a second prediction more accurate.

TABLE 3 Comparison of fixed-effect estimates and standard errors from three parametric models

Covariate	LMM	FPMM	RSMM
PSA at age 50 years	0.628 (0.068)	1.33 (0.079)	1.30 (0.075)
Age	0.073 (0.004)	0.011 (0.008)	0.023 (0.007)
Age ³	–	6.8 × 10 ⁻⁵ (1 × 10 ⁻⁵)	–
^a (Age – 12.8) ₊	–	–	0.030 (0.011)
^a (Age – 17.7) ₊	–	–	0.026 (0.012)
Gleason score = 7	–0.311 (0.168)	–0.140 (0.131)	–0.119 (0.138)
Age × Gleason score	0.033 (0.011)	0.025 (0.009)	0.024 (0.009)

^a $\alpha_+ = \alpha$ if $\alpha > 0$ and 0 otherwise.

Fractional polynomial mixed model

The fractional polynomial model of degree 1 (FP1) with the lowest deviance was that with a cubic age term as a covariate, the FP2 selected involved linear age and cubic age. *Table 4* shows how the FP2 model was chosen in favour of the FP1 and linear model. We include random coefficients for both polynomial terms to allow for each individual to have their own linear and cubic effect of age on log(PSA).

The interaction between the cubic effect of age and Gleason score is omitted because it was not found to improve the model fit. The inclusion of the cubic age term indicates the inadequacy of the linear model for age. It is important to realise that the cubic term-only model was found to be the best FP1, and the addition of the linear term to it (in the FP2 model) was found to improve the fit through a hypothesis test (i.e. *Table 4*, Step 3). The interaction between age and Gleason score more than trebles the effect of age on log(PSA) for those in the higher Gleason category (from 0.011 to 0.036). If a Gleason score by cubic age interaction were included, the interpretation of this interaction coefficient would not be useful or straightforward.

The intercept [estimate of log(PSA) at age 50 years] is much larger here than in the LMM. In *Figure 4* we see that the FPMs allow a more versatile range of fits to the simulated PSA profiles compared with LMMs. The cubic term allows for a steeper gradient in PSA for older people. Overall, there is a positive effect of age on log(PSA), which concurs with the LMM.

TABLE 4 Results of the fractional polynomial testing procedure

Model	Deviance	Power(s)	Step	Comparison	Deviance difference	p-value
FP2	1067.27	1, 3	–	–	–	–
Null	4543.64	–	1	FP2 vs. null	3476.37	<0.001
Linear	1415.82	1	2	FP2 vs. linear	348.55	<0.001
FP1	1351.46	3	3	FP2 vs. FP1	284.19	<0.001

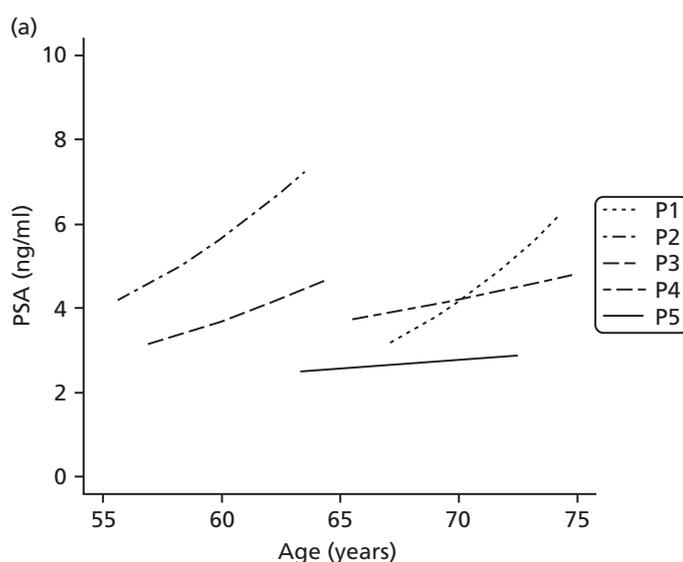


FIGURE 4 Fits to the five simulated PSA profiles. (a) LMM; (b) FPM; (c) RSM; and (d) PACE. (continued)

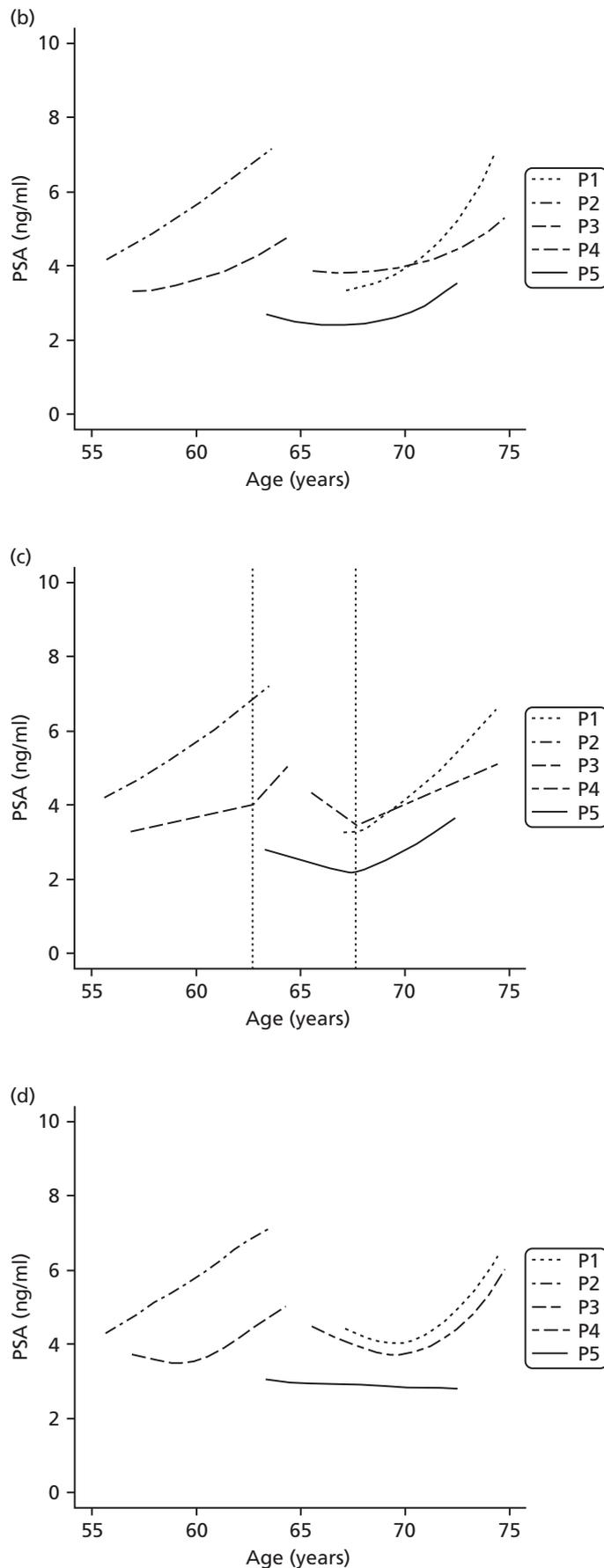


FIGURE 4 Fits to the five simulated PSA profiles. (a) LMM; (b) FPMM; (c) RSMM; and (d) PACE.

Regression spline mixed model

Here, $K = 2$ knots were selected, based on AIC, to gain modelling flexibility while avoiding potential overfitting of these PSA profiles. Given that no evidence of natural 'jumps' of PSA occur at any particular age for all men, these two knots were globally placed at the 33.3 and 66.7 centiles of the distribution of age. All interactions between spline terms and Gleason score were added initially but then removed owing to no model improvement.

Table 3 shows that the linear relationship between age and PSA becomes stronger as a man gets older. Indeed, the strength is doubled at 63 years and trebled at 68 years, compared with the trend for a man aged < 63 years. As can be seen in *Figure 4*, the fits are allowed to break from a linear trend to capture some features missed by the LMM.

Functional principal components analysis

The first two principal components of the PSA data accounted for 92% of the variation in PSA. In *Figure 4* the serial PSAs for each man are fitted with smooth curves, which are deviations from the overall mean function $\mu(t)$ displayed in *Figure 5*. These individual curves are not restricted by a parametric shape. The mean curve $\mu(t)$ suggests a much less extreme trend than each of the other approaches. PSA is estimated to rise slowly up to the age of 70 years, at which point a sudden change of slope occurs with PSA rising at a faster rate. This fit appears similar to a linear regression spline fit with one knot at 70. The underlying model gives this fit far more flexibility, and, of all methods, it estimates the slowest trend for PSA.

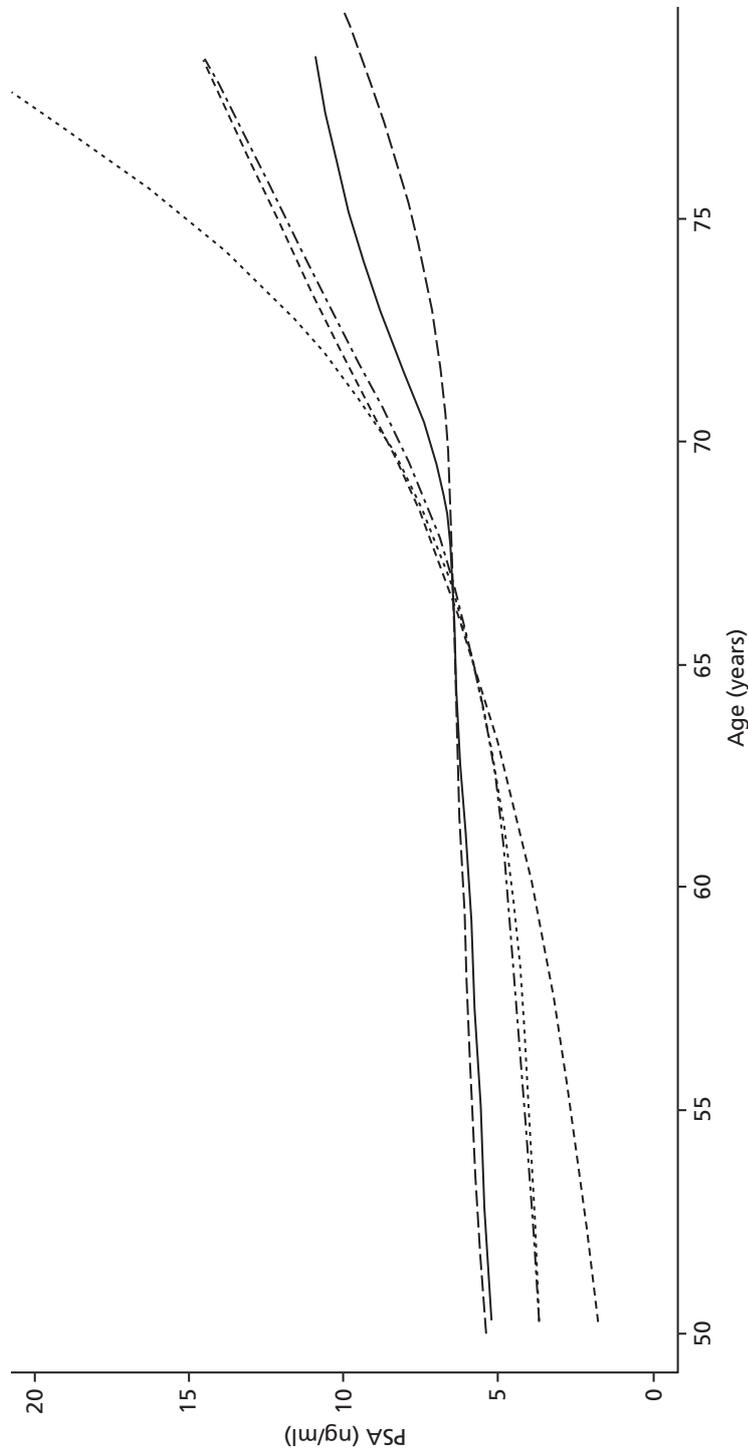


FIGURE 5 Mean PSA over age (solid line) with fits using LMM (dashed), FPMM (dotted), RSMM (dash dotted) and PACE (long dashed).

Choosing a model for prostate-specific antigen reference ranges

Describing the mean pattern of prostate-specific antigen change

Tables 5 and 6 summarise the difference in fitted and observed PSA and the RMSE, respectively, across age categories. In each age category, PACE is superior for describing these data. Apart from in the right tail (at ages > 71 years), the difference between the observed PSA and that fitted by the model is smaller for PACE than for the other methods. If the sole interest here was in representing PSA trend for this group of men on AM, then PACE would be the best choice. Among the other methods, from ages 50–60 years, the LMM performs quite badly. As can be seen in Table 5, the fitted PSA values are obviously not high enough on average. Over half of the total observations of PSA were taken from men aged 61–70 years, and this range sees comparable performance of the parametric methods. Figure 5 serves to illustrate the benefit of using RSMM instead of FPMM in this example. Up to the age of roughly 70 years, both have a similar mean estimate. The FPMM then continues on this path because it uses a global polynomial basis for model fitting, which leads to severe overestimation of mean PSA in the right tail of age. However, the regression spline fit has the flexibility to change at the second knot (i.e. at age 68 years) and this change then becomes apparent with a flatter fit of the regression spline in the right tail. Outlying PSA measurements in older age affect only the spline-based estimate of the age effect in that region, whereas with FPMM they have a global effect on the estimate.

From Table 7, the FPMM and RSMM have lower AIC compared with the LMM, but only marginally more variation is explained by the two more complex models. The RMSE is lowest using PACE followed by the RSMM approach, which corroborates the results from Table 6. Based on these diagnostic elements, the RSMM is the best fitting of the three parametric models for PSA in men on AM.

TABLE 5 Fits to the observed PSA (ng/ml) by age category

Age category (years)	Number of observations	Mean LMM fitted: observed PSA (SD)	Mean FPMM fitted: observed PSA (SD)	Mean RSMM fitted: observed PSA (SD)	Mean PACE fitted: observed PSA (SD)
50–55	485	–2.9 (2.9)	–1.5 (2.9)	–1.4 (2.9)	0.66 (2.5)
56–60	1222	–2.3 (3.9)	–1.5 (3.9)	–1.3 (3.9)	0.18 (3.6)
61–65	2422	–1.1 (3.9)	–0.88 (3.9)	–0.88 (3.9)	0.08 (3.7)
66–70	2320	0.86 (4.2)	0.80 (4.2)	0.52 (4.2)	–0.09 (3.8)
71+	989	2.0 (5.9)	3.4 (6.3)	1.6 (5.9)	–0.52 (4.2)
All	7438	–0.39 (4.5)	0.80 (4.5)	–0.24 (4.4)	0.0005 (3.7)

TABLE 6 Root-mean-square error of PSA (ng/ml) by age category

Age category (years)	Number of observations	LMM RMSE (SD)	FPMM RMSE (SD)	RSMM RMSE (SD)	PACE RMSE (SD)
50–55	485	2.98 (2.9)	2.08 (2.3)	2.11 (2.4)	1.75 (1.9)
56–60	1222	2.72 (3.6)	2.39 (3.3)	2.39 (3.3)	2.24 (2.8)
61–65	2422	2.63 (3.0)	2.63 (2.9)	2.63 (2.9)	2.30 (2.9)
66–70	2320	3.08 (3.0)	3.02 (3.0)	2.93 (3.1)	2.25 (3.1)
71+	989	4.85 (3.9)	5.53 (4.1)	4.66 (3.95)	2.48 (3.4)
All	7438	3.11 (3.3)	3.06 (3.3)	2.92 (3.3)	2.26 (3.0)

TABLE 7 Model diagnostics and RMSE (in ng/ml) for the parametric approaches (no AIC is available from the PACE method)

Model performance measure	LMM	FPMM	RSMM
AIC	1412	1084	994
σ_e^2	0.045	0.043	0.042
RMSE of conditional prediction (ng/ml)	2.15	2.19	2.10

Table 3 compares the estimates of the associations between Gleason score and PSA trajectory from the three parametric models. There are some similarities in the effect of Gleason score and its interaction with age; however, the intercept is much lower in the LMM than in either of the other two methods (which have very similar estimates). In the LMM, with a constant trend over time, the intercept must be low to allow for the model to fit the higher PSA values in the right tail of age. We can see in *Table 5* that this leads to an underestimate of PSA when age is between 50 and 60 years. For the same reason (i.e. to allow those with higher Gleason score to have higher PSA in later age) the LMM underestimates the association between Gleason score and PSA at age 50 years and overestimates the association between Gleason score and PSA slope, compared with the FPMM and RSMM. The RSMM or FPMM methods are preferred here for estimating the effect of Gleason score on PSA change over time.

Predicting prostate-specific antigen for a future patient

Here a parametric model is preferred, given that, by using methods that condition on new observations of the response,^{98,100} we can allow the parametric models to become individualised to newly diagnosed men on monitoring. Without a 'factor loading' ξ_{i1} , a new individual i would simply be predicted to lie on the mean PACE trajectory given his age.

Table 7 presents the RMSE of fitted PSA for each of the 512 men conditioned on their initial PSA. Here we can see that the conditioning improves on the parametric methods such that they are better at fitting these data than PACE. Between the three approaches, the RSMM is still preferred. Surprisingly, the LMM performs slightly better than the FPMM when fitted values are conditioned on the first observation of PSA. Regression splines offer the optimum performance among parametric methods overall (with the lowest AIC and RMSE for both unconditional and conditional predictions) and also perform well in all segments of age, as shown in *Table 5* and *Table 6*. Thus, using regression splines for prediction of PSA in men in the future seems appropriate. This reflects the changing trajectory of PSA over age, given that the RSMM allows for three separate effects of age on PSA.

Summary

This chapter has compared three parametric multilevel models (LMM, FPMM, RSMM) and a non-parametric smoothing method in order to choose a model for longitudinal PSARRs. Using PACE gave the best descriptive fit to the data initially, but, using methods that condition on initial PSA, the parametric methods displayed lower RMSE. As a predictive model, the linear RSMM, with two knots placed at the 33.3 and 66.7 centiles of age, offers the best model for PSA in this population.

Chapter 4 Testing the accuracy of the prostate-specific antigen model in four separate cohorts

Aims and background

Here, we test whether or not the model for PSA presented in *Chapter 3* is suitable for general application, by using it to predict circulating PSA levels in four external cohorts of men on AM/AS. We summarise the predictions for each cohort comparing the ProtecT model with a model based on all the data from each cohort in turn. If the ProtecT model is found to describe age-related PSA changes in other cohorts, then each PSA measured for a man in any clinic can be compared with the value expected with age-related change.

Methods

Study populations

The RMH AS cohort is an ongoing study into the impact of initial conservative management of clinically localised PCa.¹⁰¹ Data from 499 men, comprising 9472 PSA tests along with Gleason score and several other clinical covariates, were available. PSA test results were obtained between 1999 and 2012. The study eligibility criteria were baseline PSA < 15 ng/ml, Gleason score of $\leq 3 + 4$ and percentage of positive biopsy cores $\leq 50\%$. In most men, diagnosis was based on a raised PSA and subsequent positive biopsy, so they represent a modern AS cohort. Men on AS were followed up with PSA tests every 3–4 months in the first 2 years and every 6 months thereafter. Biochemical progression was defined as PSA velocity > 1 ng/ml/year, whereas histological progression on rebiopsy was defined as having a Gleason score of $4 + 3$ or > 50% positive biopsy cores. Clinical outcomes of metastases and PCSM were collected.

The Johns Hopkins (JH) AS programme began recruitment of men with clinically localised PCa in 1995.⁸² Men were eligible if they had a Gleason score of $\leq 3 + 3$, T1c, PSAD < 0.15 ng/ml/cm³, two or fewer positive biopsy cores and maximum involvement of 50% per core. Data from 961 men, comprising 9993 PSA test results performed between 1993 and 2012, along with diagnostic Gleason score and several other clinical covariates, were available. Radical treatment was recommended once men no longer met the eligibility criteria described above. Clinical outcomes of all-cause mortality and PCSM were included in the data.

The SPCG4 data contained 290 men with 2987 PSA tests.¹² These men were randomised to WW as part of the SPCG4 RCT comparing WW and radical prostatectomy. They were diagnosed between 1989 and 1999, for the most part through clinical presentation with symptoms. To be suitable for randomisation the men were required to be < 75 years of age, with a life expectancy of > 10 years. Data were available on whether or not the men developed metastases and/or died from PCa.

Data received from the University of Connecticut Health Center (UCHC) cohort consist of 114 men with 884 PSA test results¹⁰² followed for an average of 4.7 years (SD 3.9 years). Men were diagnosed between 1989 and 1993 (before the advent of widespread PSA testing), for the most part by clinical presentation with symptoms. Fewer baseline clinical exposures were available than from other cohorts, and information was not present on whether or not men developed metastases or died from PCa.

These four cohorts come from two eras, the PSA detection era (RMH and JH) and the clinical detection era (UCHC and SPCG4). As described above, the advent of PSA screening resulted in many more men being diagnosed with PCa at an earlier stage than before. The modern cohorts also come from populations with different screening prevalence. In the USA there is widespread PSA screening, whereby men are likely to have several PSA tests through their lifetime. Thus, men diagnosed with PCa in the USA are likely to have had mostly 'normal' PSA results before the raised value, which resulted in a biopsy and subsequent diagnosis. In the UK there is no such screening programme, and the men diagnosed with PCa in RMH may have had any level of PSA before the single high PSA that led to their diagnosis. Hence, the US men are likely to be a lower-risk group with lower PSA on average than their UK counterparts. These issues, and their impact on the poorly understood natural history of PCa, need to be considered when interpreting findings from the cohorts.

The coefficients found in the ProtecT trial model were applied to data from the RMH and JH AS cohorts as well as the UCHC and SPCG4 WW cohorts. However, all data were restricted to PSA test results ≤ 50 ng/ml and to men with an initial PSA ≤ 20 ng/ml. This is to eliminate any atypical values, in terms of a modern AM/AS study. Two cohorts of men without PCa were also included to examine differences in PSA change between men with and without cancer. Data from the Baltimore Longitudinal Study of Aging (BLSA)¹⁰³ contained repeated 5012 PSA measurements of 1032 men. A model for PSA change in 1432 men without cancer from the Krimpen study,¹⁰⁴ a large prospective community-based study in the Netherlands, has appeared elsewhere.³⁵

Measuring performance

First, to investigate the behaviour of PSA between cohorts, we fit a simple random intercept and slope multilevel model to $\log(\text{PSA})$ in each of these four cohorts:

$$\log(\text{PSA}_{ij}) = \beta_0 + u_{0i} + (\beta_1 + u_{1i})\text{age}_{ij} + \varepsilon_{ij}, \quad (4)$$

where u_{0i} is the random intercept, which allows each man to have their own adjustment to the intercept β_0 [i.e. the average value of $\log(\text{PSA})$ at age 50 years] and u_{1i} is the random slope, which allows each man to have their own adjustment to the slope β_1 (i.e. the average change in $\log(\text{PSA})$ per year increase in age). This is carried out to compare the age-related change in PSA for men with and without PCa.

The coefficients for our model of PSA change have been previously estimated using a cohort of 512 men with clinically localised PCa participating in the ProtecT study.⁹³ In the present analysis, the model is used to predict PSA in each of four external cohorts. We measure the accuracy of the predictions using the average difference between observed and predicted PSA value per PSA test:

$$\frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \sqrt{(\text{PSA}_{ij} - \widehat{\text{PSA}}_{ij})^2}}{N}, \quad (5)$$

where PSA_{ij} is the observed PSA test result for person $i = 1, \dots, n$ measured at time $j = 1, \dots, n_i$ and $\widehat{\text{PSA}}_{ij}$ is the predicted PSA test result for person i at time j . The number of PSA tests can be different from person to person, and the total number of PSA tests in the cohort is N . The average absolute difference between observed and predicted PSA will always be positive, with a value of zero if the model predicted PSA perfectly.

For each of the four cohorts, predictions are made using (a) a model with coefficients derived from the external data themselves; (b) the ProtecT model coefficients; and (c) the ProtecT model updated using the first three PSA values for each man in the external cohort.^{97,98} Prediction (a) gives the hypothetical upper limit of performance but is not clinically useful, because the model coefficients cannot be estimated until all the PSA measurements have been taken over the duration of AM. Predictions (b) and (c) indicate what could be

achieved in clinical practice, as they apply coefficients estimated using the ProtecT cohort to the other data sets, and so can be applied each time a new measure becomes available for a man on monitoring. To calculate the coverage of the model in predicting PSA, we check whether or not a 95% prediction interval (calculated using unconditional standard errors) from the ProtecT model contains the corresponding observed value of PSA. We measure performance of the models further by tabulating the model failures, which we define as absolute difference between predicted and observed PSA > 5 ng/ml and model successes, defined as predicted PSA within 2 ng/ml of observed PSA. These are tabulated to obtain the proportion of test level failures/successes (i.e. for how many PSA test results does the model fail/succeed) and subject level failures/successes (i.e. for how many men does the model fail/succeed on average across all their PSA test results). These cut-offs were chosen to reflect what we believe to be clinically significant ranges.

Results

Age-related prostate-specific antigen change

Men on AM/AS and men without PCa have similar age-related PSA change (*Table 8*). For example, the PSA change per year is very similar in the Krimpen, BLSA, RMH and JH cohorts. *Figure 6* shows the predicted average pattern of change if each cohort had an average PSA of 2 ng/ml at age 50 years. This hypothetical graph shows the similarities of the four modern cohorts involving men with or without PCa. However, the results from the multilevel models suggest that men without cancer have much lower average PSA values at age 50 years. In *Figure 7* we see that the estimated average PSA at age 50 years is much lower in the Krimpen and BLSA cohorts.

TABLE 8 Coefficients from simple models of PSA change across six cohorts

Variable	Men with no PCa		Men with PCa			
	Krimpen	BLSA	RMH	JH	SPCG4	UCHC
Average PSA at age 50 years	0.72	0.65	2.48	2.07	1.38	1.23
Percentage change in PSA per year in age (%)	4	4	5	4	11	7

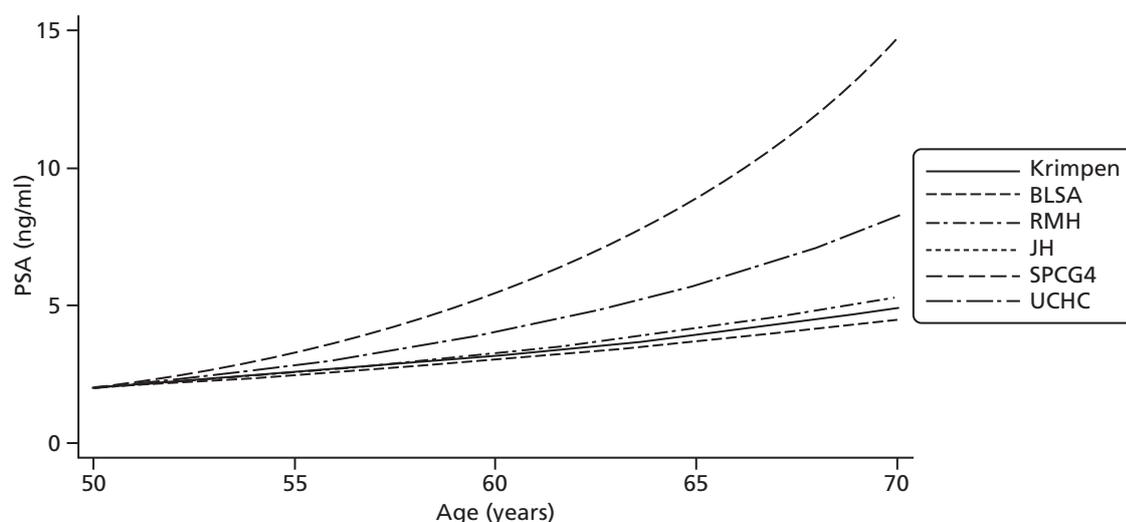


FIGURE 6 Hypothetical PSA change in the cohorts if average PSA at 50 years was 2 ng/ml in each.

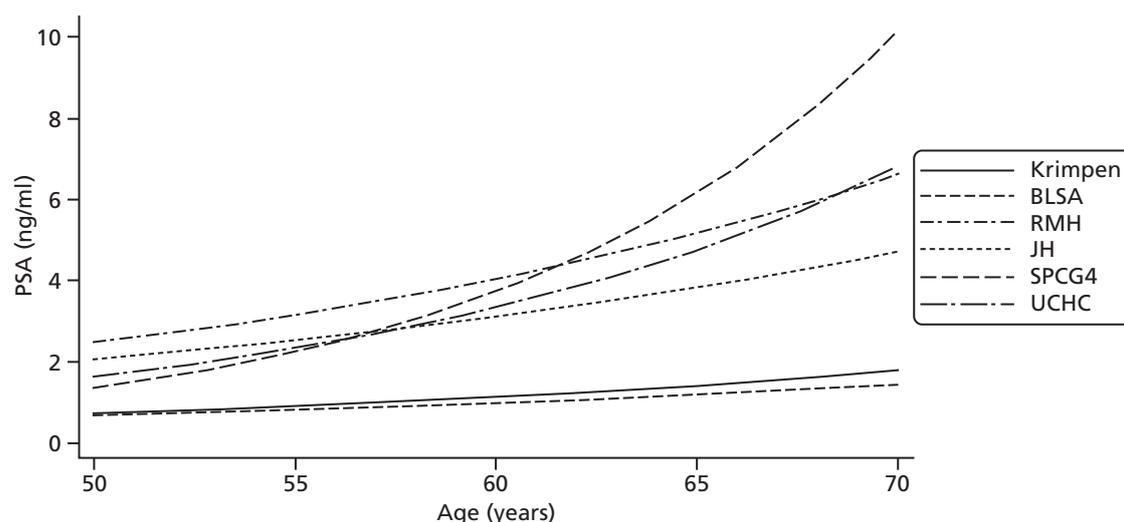


FIGURE 7 Estimated PSA change in the cohorts.

Screen-detected cohorts

The model updated with the initial three PSA results is, on average, 2 ng/ml away from the true PSA for RMH and 1.8 ng/ml away from the true PSA for JH (Table 9). Models fitted using all the PSA data from these AS cohorts lead to an average difference between observed and predicted PSA of 1.1 ng/ml for RMH and 0.82 ng/ml for JH. This indicates that the best prediction that can be achieved after collecting all PSA data from this cohort. Furthermore, the coverage of the model (i.e. whether the observed value lies within the 95% prediction interval) is close to the nominal value of 95%, at 98% for RMH and 97% for JH.

TABLE 9 Average absolute difference between observed and predicted PSA and coverage of observed PSA using the protect model on four separate cohorts

Cohort	Hypothetical, retrospective model using all PSA data	Fixed prediction using ProtecT model	Prediction updated using observed first, second and third PSA values
RMH			
Average absolute difference between observed and predicted PSA, ng/ml (SD)	1.1 (1.5)	3.3 (3.0)	2.0 (2.9)
Coverage	0.999	0.93	0.98
JH			
Average absolute difference between observed and predicted PSA, ng/ml (SD)	0.82 (1.3)	3.1 (2.8)	1.8 (2.5)
Coverage	0.997	0.86	0.97
SPCG4			
Average absolute difference between observed and predicted PSA, ng/ml (SD)	1.7 (2.5)	7.4 (7.6)	4.6 (6.1)
Coverage	0.999	0.75	0.95
UCHC			
Average absolute difference between observed and predicted PSA, ng/ml (SD)	1.3 (2.0)	5.6 (5.7)	3.7 (5.6)
Coverage	0.996	0.77	0.96

Updating the model using the initial three PSA values has improved the accuracy of the ProtecT model. In both RMH and JH the difference between the observed and predicted PSA values reduces by 1.3 ng/ml per PSA test when using the updated model instead of the standard one.

Clinically detected cohorts

The ProtecT model achieves less accurate predictions in the older cohorts of SPCG4 and UCHC (see *Table 9*). Although an optimal model fitted using all the SPCG4 data gives values of PSA within an average of 1.7 ng/ml from the observed PSA, the ProtecT predictions differ by 4.6 ng/ml on average. Similarly, in the UCHC cohort, the ProtecT model leads to predictions that differ by 3.7 ng/ml from the observed PSA on average, whereas, for an optimal model fitted using all the UCHC data, the average difference between observed and predicted PSA is 1.3 ng/ml. However, coverage was 95% for SPCG4 and 96% for UCHC. The results from both these cohorts further demonstrate the improvements gained using the model updated with the initial three PSAs, with the average difference per test reducing by 2.8 ng/ml in SPCG4 and 1.9 ng/ml in UCHC.

Comparison across cohorts

A clear dichotomy between the cohorts is found in *Table 10*. Roughly 3–4% of men have an average absolute difference between predicted and observed PSA of > 5 ng/ml using the external model in the modern AS cohorts. However, this rises to 14% and 30% for the UCHC and SPCG4 cohorts, respectively. A similar divide is apparent in the percentage of PSA tests where the model fails. This is 7–8% for the modern AS cohorts but 20% and 25% in the UCHC and SPCG4 cohorts, respectively. The reverse of these results is also true, in that the amount of model successes shows up a dichotomy between the cohorts (*Table 11*). Between 67% and 79% of men in the modern AS cohorts have an average difference between predicted and observed PSA of < 2 ng/ml across all their PSA test results. The two older cohorts have a lower proportion of men, between 39% and 51%, whose average absolute difference between predicted and observed PSA is within 2 ng/ml. The percentage of individual predicted PSAs within 2 ng/ml of the actual test results is 70–73% in the modern cohorts but 55% and 47% in UCHC and SPCG4 respectively.

TABLE 10 Model failure by study using prediction conditioned on first, second and third PSA

Study	Average absolute difference between observed and predicted PSA per individual > 5 ng/ml (%)	Average absolute difference between observed and predicted PSA per PSA test > 5 ng/ml (%)
RMH	4	8
JH	3	7
SPCG4	30	25
UCHC	14	20

TABLE 11 Model success by study using prediction conditioned on first, second and third PSA

Study	Average difference between observed and predicted PSA per individual < 2 ng/ml (%)	Average difference between observed and predicted PSA per PSA test < 2 ng/ml (%)
RMH	67	70
JH	79	73
SPCG4	39	47
UCHC	51	55

Summary

The ProtecT trial model leads to a useful prediction of PSA in AS cohorts, with up to 79% of men having an average difference between predicted and observed PSA of < 2 ng/ml. Predictions from the model were less accurate in the two older cohorts, made up of men probably at a later stage of disease and with higher PSA values. The model predicts with an average difference between predicted and observed PSA of roughly 2 ng/ml in both screen-detected cohorts and roughly 4 ng/ml in the symptom-presenting cohorts. However, using a hypothetical model based on all the collected data from these cohorts, the average difference between observed and predicted PSA is still between 0.8 ng/ml and 1.7 ng/ml. The dichotomy of performance was evident in the number of men not well described by the model (i.e. men whose average difference between observed and predicted PSA across their PSA test results was > 5 ng/ml). In the AS cohorts there were 3–4% of men with an average difference between observed and predicted PSA > 5 ng/ml in the prediction model, whereas 14–30% had at least this average difference between observed and predicted PSA in the older cohorts. In predicting PSAs across all models, the benefit of using an updated model resulted in reductions in difference between observed and predicted PSA for each cohort.

A large number of data were available for model development and validation. The model was developed using data from over 7000 PSA tests, and we have attempted to predict over 15,000 PSA test results in further cohorts using this model. Our data come from both sides of the Atlantic and traverse two eras of PCa detection: the symptomatic-presenting man from the early 1990s and the screen-detected man of the 2000s. The model performs well in both the modern US and European cohorts even though it was developed on a UK cohort. This suggests that, although PSA at diagnosis is likely to be lower in the USA, PSA change is similar. The model performs less well in predicting PSA for men in the older cohorts, but some of these men with clinically detected disease could be seeing PSA changes beyond those expected. These rapid rises of PSA would not be predicted by the model and could potentially be above a 95% reference range. In some ways, the lack of performance in these cohorts can be viewed as evidence for the use of PSARRs in AM.

Chapter 5 Comparison of prostate-specific antigen markers used to alert men to worsening prostate cancer

Aims and background

This part of the project involved:

- investigating the prediction of PCa outcomes
- validating the ability of a PSARR to predict PCa outcomes
- testing the PSARR method against the commonly used methods of PSADT and PSAv.

Chapter 4 above has shown that the model of age-related PSA change offers reasonably accurate predictions of future PSA levels in modern monitoring cohorts. In this chapter, we compare PSARRs with PSADT and PSAv in retrospective analyses of three historical cohorts where several measures of PCa progression are available.

Methods

Prostate-specific antigen markers

Prostate-specific antigen doubling time is the most popular PSA kinetic measure used to identify those patients on AM who need clinical review. PSADT is calculated from a model of $\log(\text{PSA})$ against time:

$$\log(\text{PSA}) = \alpha + \beta \times \text{time}. \quad (6)$$

The doubling time is then calculated as $\log(2)/\beta$ and is interpreted as the estimated time for a man's PSA to double. Thus, low values of doubling time coincide with rapidly rising PSA and are a cause for concern. Conversely, high values correspond to a steady trend of PSA, and those patients above a threshold of PSADT are advised to remain on monitoring. For a patient's doubling time to be calculated using the regression method, at least three PSA test results are required. We consider a PSADT of < 3 years to be a marker for rapidly rising PSA.^{79,84,86,89}

An alternative to PSADT is PSAv. This explicitly measures the rate of change in PSA over time calculated as the slope term (β) from a model of PSA against time on monitoring:

$$\text{PSA} = \alpha + \beta \times \text{time}. \quad (7)$$

Given that PSAv is the rate of increase of PSA, low positive values refer to steady PSA change and high PSAv equates to sharply increasing PSA. Several studies have used a threshold of $\text{PSAv} > 1 \text{ ng/ml/year}$ ^{22,71,100} as a trigger for clinical review during monitoring. For this paper, we have calculated each man's PSAv at each visit beyond their third (calculation of PSAv requires at least three observations of PSA). We have used $\text{PSAv} > 1 \text{ ng/ml/year}$ as a marker for rapidly rising PSA.

Prostate-specific antigen reference ranges have been established as alternatives to these two PSA kinetic measures.³³ A model for such reference ranges has been previously developed⁹³ on a cohort of 512 men enrolled in the AM arm of the ProtecT trial.¹⁵ For example, a 95% reference range, above which the fastest increasing 5% of men with clinically localised PCa lie, is constructed as:

$$\widehat{PSA} + 1.68\sqrt{\widehat{var}(PSA)}. \quad (8)$$

where \widehat{PSA} and $\widehat{var}(PSA)$ are found using the ProtecT trial model described in *Chapter 3*. If subsequent PSA test results fall above this reference range, this implies that the patient's PSA growth is above what would be expected as a result of normal ageing, such that a recommendation for clinical review could be warranted (e.g. *Figure 8*). Using the reference range model, we calculate whether or not a man's PSA result is inside the reference range at each visit; if it falls outside the reference range, this suggests an increasing PSA beyond expected normal changes and that further clinical review may be necessary.

Study populations

The PSA markers were tested on the RMH AS cohort and the SPCG4 and UCHC WW cohorts described in *Chapter 4, Methods, Study populations* (see also *Table 24*). The data were again restricted to remove any single PSA test result > 50 ng/ml and any man with a PSA at diagnosis of > 20 ng/ml. Hence, the sample sizes differ (*Table 12* and see *Table 25*).

Outcomes

All-cause mortality data from UCHC are not used in this analysis to validate reference ranges, because they probably contain many non-PCa-related deaths. Although the data from RMH include PCSM, no such deaths are reported. To exploit these cohorts fully, we introduce outcomes pertaining to a PSA milestone, namely whether or not a man reaches a PSA of 20 ng/ml during follow-up. The outcomes used for validation, along with the proportion of men experiencing these outcomes, are given in *Table 12*. Given that it is not clinically useful to have an alert (through PSADT, PSA_v or PSARRs) close to, say, PCSM, we have only included those alerts occurring at least 5 years prior to PCSM or metastases and 2 years prior to reaching a PSA of 20 ng/ml.

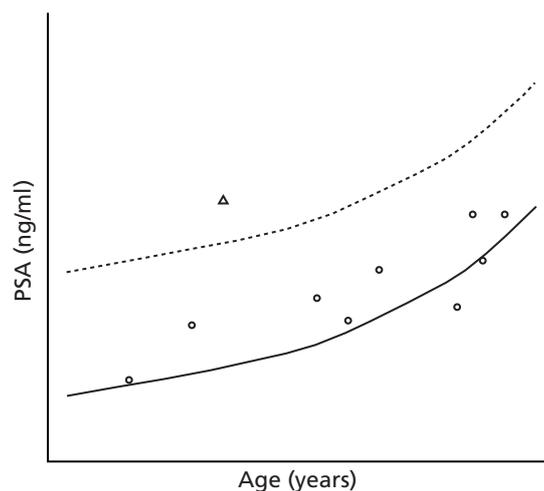


FIGURE 8 An example of a PSARR. The PSA result represented by a triangle would lead to a recommendation for clinical review.

TABLE 12 Summary of PCa outcomes used for validation in each cohort, with the percentage positive for each outcome during cohort follow-up

Cohort	Gleason score of 3 + 3	Gleason score of 3 + 4/4 + 3
SPCG4		
Sample size	180	88
Metastases	44 (26%)	28 (40%)
PCSM	28 (15%)	27 (33%)
RMH		
Sample size	454	45
PSA > 20 ng/ml	47 (10%)	8 (18%)
UCHC		
Sample size	86	32
PSA > 20 ng/ml	18 (21%)	12 (38%)

Statistical analysis

Statistical analyses were performed in Stata 12.⁷⁰ We report results solely for men with Gleason score of 3 + 3, because this low-risk group represents the vast majority of modern AM cohorts.¹⁷ We first compared 90%, 95% and 99% PSARRs to choose the optimal cut-off for rapidly rising PSA using the area under the receiver operator characteristic (ROC) curve statistic for onset of metastases and PCSM in SPCG4. Cox proportional hazards (Cox PH) models¹⁰⁵ were used to estimate the association between markers and time-to-event outcomes. Time zero was set to be the date of beginning of monitoring, whereas date of an event or end of monitoring was the time at risk. For each cohort we fit a base model controlling for T-stage (where applicable), baseline PSA and baseline age. We report the HRs for each PSA marker and outcome, together with their 95% CIs. Finally, we compare PSADT, PSA_v and PSARR using the difference of deviances from the Cox models using each of these in turn. A better model will have a lower deviance (see *Table 16*).

The ability of each PSA marker to discriminate between higher- and lower-risk men is calculated using Harrell's concordance statistic (or c-statistic) from each Cox PH model.^{105,106} This is equivalent to an area under the ROC curve for time-to-event responses, such as those of interest here. The c-statistic estimates the probability that, of two randomly chosen men, the man with the higher-predicted probability of an outcome will reach an outcome before the patient with the lower-risk score. By adding each PSA marker in turn to the baseline models, we can compare the change caused by the marker, by looking at the difference in c-statistic from base model to marker model.

Results

For onset of metastases, using 95%, 97.5% and 99.5% reference ranges led to an area under the curve (AUC) of 0.58, 0.59 and 0.55, respectively. For the PCSM outcome, using 95%, 97.5% and 99.5% reference ranges resulted in AUCs of 0.59, 0.56 and 0.54 respectively. Given these results, we chose to proceed using the 95% reference ranges.

Metastases

Using an alert based on PSARR alone to recommend clinical review would miss 15 out of 30 future metastatic cancers (sensitivity 50%, 95% CI 31% to 69%) while causing 38 unnecessary rebiopsies to be performed in the 107 men with non-life-threatening disease (specificity 65%, 95% CI 55% to 74%). The PSADT < 3-year marker would lead to 13 missed metastatic cancers out of 30 (sensitivity 57%, 95% CI 37% to 75%; *Table 13*) and would unnecessarily biopsy 42 of the 107 men with disease that would not

progress to metastases during the follow-up of the cohort (specificity 61%, 95% CI 51% to 70%). Finally, the PSAv > 1 ng/ml/year marker alone would miss just eight of the 30 future metastatic cancers (sensitivity 73%, 95% CI 54% to 88%) but would cause 70 unnecessary biopsies (specificity 35%, 95% CI 26% to 45%).

Initial age, PSA and T-stage were not useful to identify those men developing metastases (*c*-statistic = 0.58, see *Table 15*). There was weak evidence of positive associations between the PSADT marker (HR 2.1, 95% CI 0.97 to 4.4) and PSARR (HR 2.4, 95% CI 0.9 to 6.7) and development of metastases (see *Table 15*), and no evidence that men alerted by the PSAv > 1 ng/ml/year marker were more likely to develop metastases in the future (HR 1.1, 95% CI 0.5 to 2.7; see *Table 15*). However, the PSA markers slightly improved the discriminative ability of the baseline model, with increases in *c*-statistic of 9% (PSADT), 7% (PSARRs) and 2% (PSAv). Comparing the PSA marker models, there was some weak evidence of a better predictive ability using doubling time over velocity (*p* = 0.06; see *Table 16*).

Prostate cancer-specific mortality

The PSARR marker led to 56% sensitivity (95% CI 31% to 79%; *Table 14*) and 61% specificity (95% CI 52% to 70%) when used alone to determine which men would die from PCa. This would mean that 8 out of 18 men with lethal cancer would be missed by the PSARR, and 48 unnecessary biopsies would be performed. Using the PSADT marker alone would miss 6 out of 18 lethal cancers (sensitivity 67%, 95% CI 41% to 87%) and cause 52 unnecessary rebiopsies to be performed in the 124 men without lethal cancer (specificity 58%, 95% CI 49% to 67%). If the PSAv marker were used as a standalone recommendation for clinical review, just three of the 18 lethal cancers would be missed (sensitivity 83%, 95% CI 59% to 96%), but 83 unnecessary biopsies would be performed in the 124 men with non-life-threatening disease (specificity 33%, 95% CI 25 to 42%).

A model using baseline characteristics of PSA, T-stage and age was not useful to discriminate between those who died as a result of PCa and those who did not (*c*-statistic = 0.53; *Table 15*). Those men with a PSADT < 3 years were more likely to die from PCa (HR 3.2, 95% CI 1.1 to 9.2) and this improved the *c*-statistic from the baseline model by 0.13 (25%). There was some weak evidence for an association between PSARR (HR 3.3, 95% CI 0.9 to 12.1) and little evidence for an association between PSAv (HR 2.1, 95% CI 0.6 to 7.7) and death from PCa. There was some weak evidence that the PSADT model was a better predictor of outcome than the PSAv (*p* = 0.05; *Table 16*) model.

TABLE 13 Sensitivity and specificity comparing the PSA markers in alerting men with onset of metastases

Outcome/statistic	PSADT < 3 years		PSAv > 1 ng/ml/year		PSARR	
	Alerted	Not alerted	Alerted	Not alerted	Alerted	Not alerted
Metastases	17	13	22	8	15	15
No metastases	42	65	70	37	38	69
Sensitivity, % (95% CI)	57 (37% to 75%)		73 (54% to 88%)		50 (31% to 69%)	
Specificity, % (95% CI)	61 (51% to 70%)		35 (26% to 45%)		65 (55% to 74%)	

TABLE 14 Sensitivity and specificity comparing the PSA markers in alerting men with PCa specific mortality

Outcome/statistic	PSADT < 3 years		PSAv > 1 ng/ml/year		PSARR	
	Alerted	Not alerted	Alerted	Not alerted	Alerted	Not alerted
PCSM	12	6	15	3	10	8
No PCSM	52	72	83	41	48	76
Sensitivity, % (95% CI)	67 (41% to 87%)		83 (59% to 96%)		56 (31% to 79%)	
Specificity, % (95% CI)	58 (49% to 67%)		33 (25% to 42%)		61 (52% to 70%)	

TABLE 15 Hazard ratio and c-statistic improvements for the three PSA markers on each of the four outcomes

Outcome/statistic	PSADT < 3 years	PSAv > 1 ng/ml/year	PSARR
SPCG4 metastases			
HR (95% CI)	2.08 (0.97 to 4.44)	1.13 (0.47 to 2.71)	2.41 (0.87 to 6.70)
c-statistic from baseline model	0.58	0.58	0.58
c-statistic improvement from marker (% improvement)	0.05 (9)	0.01 (2)	0.04 (7)
SPCG4 PCSM			
HR (95% CI)	3.23 (1.13 to 9.22)	2.06 (0.55 to 7.72)	3.28 (0.89 to 12.1)
c-statistic from baseline model	0.53	0.53	0.53
c-statistic improvement from marker (% improvement)	0.13 (25)	0.06 (11)	0.11 (21)
RMH PSA > 20 ng/ml			
HR (95% CI)	2.16 (1.01 to 4.60)	10.8 (1.45 to 80.8)	1.54 (0.51 to 4.09)
c-statistic from baseline model	0.74	0.74	0.74
c-statistic improvement from marker (% improvement)	0.03 (5)	0.05 (7)	0.002 (0.3)
UCHC PSA > 20 ng/ml			
HR (95% CI)	6.00 (2.82 to 12.8)	4.31 (2.47 to 7.52)	12.0 (5.38 to 26.9)
c-statistic from baseline model	0.82	0.82	0.82
c-statistic improvement from marker (% improvement)	0.02 (3)	0.02 (3)	0.03 (4)

TABLE 16 Differences in deviance (predictive ability) between the three PSA marker models

Comparison	SPCG4 metastases	SPCG4 PCSM	RMH	UCHC
PSAv–PSADT (p -value)	3.5 (0.06)	3.9 (0.05)	-7.5 (0.006)	-10.2 (0.001)
PSAv–PSARR (p -value)	2.8 (0.10)	1.9 (0.17)	-11.1 (0.0009)	10.7 (< 0.0001)
PSADT–PSARR (p -value)	0.8 (0.38)	-2.0 (0.16)	-3.6 (0.06)	20.9 (< 0.0001)

Reaching a prostate-specific antigen of 20 ng/ml during surveillance

In the RMH cohort, baseline age, PSA and T-stage had a moderate discriminative ability in separating men who would reach a PSA of 20 ng/ml (c -statistic = 0.74; see *Table 15*). There was no evidence that those men with a PSA outside their reference ranges were more likely to reach a PSA of 20 ng/ml (HR 1.5, 95% CI 0.5 to 4.1), and including PSARR only trivially improved the c -statistic. There was some evidence that men with a PSADT < 3 years were more likely to reach a PSA of 20 ng/ml (HR 2.2, 95% CI 1.0 to 4.6); including PSADT improved the c -statistic by 0.03 (5%). Those who would have been alerted by a PSAv > 1 ng/ml/year were more likely to reach a PSA of 20 ng/ml (HR 10.8, 95% CI 1.5 to 80.8). The PSAv marker led to an increase of 0.05 (7%) in the c -statistic from the baseline model. Comparing these models, the PSAv marker model was preferred to those including PSARR ($p = 0.001$; see *Table 16*) and PSADT ($p = 0.006$), with some weak evidence that the PSADT model improved on the PSARR model ($p = 0.06$).

For the UCHC data, baseline age and PSA were highly discriminative of reaching PSA 20 ng/ml in at least 2 years, with a *c*-statistic for the baseline model of 0.82 (see *Table 15*). Men with a PSA result over their PSARR were more likely to reach 20 ng/ml (HR 12, 95% CI 5.4 to 26.9; see *Table 15*). Given that there are very few men in the Gleason score of 3 + 3 category who reach a PSA of 20 ng/ml, the CI is very imprecise. Including PSARR led to an increase in *c*-statistic of 0.03 (4%) over the baseline model. Associations were found between both PSADT (HR 6.0, 95% CI 2.8 to 12.8) and PSAv (HR 4.3, 95% CI 2.5 to 7.5) and reaching a PSA of 20 ng/ml in the UCHC cohort. The inclusion of each of these markers led to increases of 0.02 over the baseline model, or 3%. There was strong evidence that the PSARR model had a higher predictive ability than both the PSAv and PSADT models (both $p < 0.0001$; see *Table 16*).

Discussion

Comparing the ability of the three PSA markers to alert men to progression of their disease, the findings here are mixed. There is some weak evidence that PSADT offers better prediction of metastases and PCSM compared with PSAv, but there was no evidence that PSADT is to be preferred to PSARRs. For classifying patients who may reach a PSA milestone of 20 ng/ml, the evidence provided here is marginally in favour of using PSAv for a modern AS cohort in RMH and strongly in favour of PSARR in data on an older pre-PSA cohort from UCHC.

Using the PSARR marker results in good specificity (i.e. leads to fewer unnecessary biopsies) in the SPCG4 men. However, this marker also misses some lethal cancers in the SPCG4 cohort (lower sensitivity). The PSAv marker has the reverse effect, with few lethal cancers missed but many unnecessary biopsies performed.

Using baseline characteristics of age, PSA and T-stage was not useful to discriminate men who go on to experience either metastases (*c*-statistic = 0.58) or PCSM (*c*-statistic = 0.53). PSADT and PSARRs increased the *c*-statistic of the PCSM model by 0.13 (21%) and 0.11 (25%). An important finding of this analysis is that using baseline age, PSA and T-stage can discriminate those men who are likely to reach a PSA of 20 ng/ml (*c*-statistics of 0.74 for RMH and 0.82 for UCHC). For this outcome, there were only marginal improvements to these *c*-statistics by any of the three PSA markers.

Using the PSARR model to track men on monitoring has several advantages compared with PSA kinetics. First, it has been shown that alerts based on PSA kinetic measures are heavily associated with the number of PSA tests.³³ For instance, one high PSA result following four unremarkable tests will lead to a small value of PSADT (and possible alert to clinical review), whereas this measure would be relatively unaffected if the high result followed 20 unremarkable PSA tests. Using the PSARR is less sensitive to the number of test results available. Second, PSA modelling is based on empirical data on the natural history of age-related PSA increase in men with clinically localised PCa, rather than a mathematical model. This model is, therefore, more robust to PSA changes (i.e. has better ability to spot an abnormal test result) than using PSA kinetic estimates obtained from a single series of PSA data. Third, using one model, such as the ProtecT trial model proposed here, to predict PSA avoids the problem of inconsistent calculation of PSA kinetics. Fourth, the calculations are all done at the outset, with no further calculations required of the clinician as further PSA results are considered.

Outcomes from the SPCG4 cohort were useful for comparison, because the number of events was relatively large and many serial measurements of PSA were taken some time before these outcomes occurred. However, in the other cohorts, a lack of clinical outcomes, such as PCSM or metastases (i.e. UCHC) or a lack of events (i.e. RMH), diminishes our ability to validate the PSARR method or to compare it with PSA dynamics. We created outcomes involving PSA milestones, which may be clinically useful. However, stronger conclusions would be drawn using PCSM. Given that men enrolling in AM are the lowest-risk group, it is unlikely that many such events will be observed until many more years of follow-up pass or many more men are enrolled in such studies.

Chapter 6 Qualitative study: clinicians' and men's views on current active surveillance practices

Introduction

Recent systematic reviews have found that there is little evidence or expert consensus over the most effective clinical practice for AS or AM of PCa.^{10,16,17} Repeat measures of PSA are usually used, but clinical practice varies in terms of the triggers for clinical review of men who have opted for AM/AS.¹⁸

The previous chapters have explained how the study team used repeat measures of PSA levels in large numbers of men with and without PCa to calculate age-related reference ranges. It was envisaged that making these reference ranges available to men when discussing blood PSA results may improve the AM/AS process by (1) providing reassurance if a man's PSA levels are within the limits expected as a result of ageing, or (2) triggering a clinical review if reference ranges indicate changes in PSA levels exceed the limit expected as a result of ageing (and thus point towards possible progression of PCa). The AMS is a software tool developed specifically for clinicians to present these reference ranges to men on AM/AS when discussing PSA test results in clinic.

The present and following chapters present the main findings to emerge from the qualitative study. All clinicians and men participating in this study referred to the intervention as AS (not AM). The term 'AS' will, therefore, be used throughout this and the next chapter.

The overall aim of the qualitative study was to investigate clinicians' and patients' views surrounding the acceptability and feasibility of using the AMS for AS of clinically localised PCa. In order to understand how the AMS may be incorporated into practice, there was an initial need to understand current AS practices, particularly as so little research has investigated experiences of providing or undergoing AS. An understanding of current practice was crucial given that the management of early-stage PCa – particularly AS – is a rapidly evolving field with little consensus in published evidence (as outlined in *Chapter 2*). In light of these considerations, specific objectives of the qualitative study were to:

1. investigate current practices within AS for clinically localised PCa
2. explore clinicians' and patients' views and experiences surrounding current AS practices (particularly their views on triggers for considering radical treatment, and current AS protocols)
3. explore clinicians' and patients' views on how the AMS compares with current practice
4. investigate which presentations of the AMS were most acceptable to clinicians and men.

The qualitative research is presented in two chapters. This chapter presents findings relating to current AS practices (objectives 1 and 2), including some comparison with the AMS, and the next chapter (see *Chapter 7*) focuses on views about the AMS in more depth (objectives 3 and 4).

Methods

Design

Semi-structured, face-to-face interviews were conducted with clinicians, nurses and patients from four hospitals, each in a separate acute trust. Acute trusts (study sites) were purposively selected on the basis of having an established, fully running AS programme. One of the four sites was involved in the ProtecT study, but clinicians interviewed at this site had no direct involvement in the study.

Sampling and recruitment of clinicians/nurses

Clinicians were purposefully selected based on their roles and non-involvement in the ProtecT study. Clinicians who led routine AS appointments were prioritised, but staff with more peripheral involvement in AS patient care were also interviewed (e.g. nurse specialists).

Principal investigators at each site suggested appropriate potential participants and helped to organise introductory meetings in two of the four sites. These meetings allowed the research team to gain preliminary insights into the structure and running of the clinics and to recruit clinicians. A degree of snowball sampling, where interview participants suggested potentially eligible colleagues, occurred at three of the four sites. Clinicians/nurses were invited to take part in the study through a letter from the research team and a study information sheet. These information sheets were either sent to the hospitals or distributed at the introductory meetings. There was potential to continue recruitment across some of the sites, but sampling ceased once data saturation had been achieved; this was defined as the point at which four consecutive interviews generated no new themes.

Sampling and recruitment of men on active surveillance

Recruitment of men on AS took place across three of the four sites. The site involved in the ProtecT trial was excluded from this phase of the study owing to concerns that men's care could have been influenced by the ProtecT study protocol for AM if patients had received care from clinicians involved in the ProtecT study. Clinical teams from each site identified eligible men in line with inclusion and exclusion criteria stated in *Box 1*. Recruitment packs were distributed to eligible men via post or within clinics. Each pack contained a letter from the researcher (LR), a reply slip, a study information sheet and a pre-paid return envelope. Men interested in participating in the study were asked to contact the researcher (LR) directly.

BOX 1 Inclusion and exclusion criteria for selecting men

Inclusion criteria

- Able to give informed written consent to participate.
- Gleason score of 6.
- PSA level at time of diagnosis < 10 ng/ml.
- Receiving AS for PCa.
- At least 12 months post diagnosis.
- Judged by clinician to show no evidence of disease progression.

Exclusion criteria

- ≥ 70 years.
- Concomitant or past malignancies (other than a small treated skin cancer).
- Prostate cancer has been diagnosed less than 12 months previously.
- Gleason score of 7 or above.
- PSA result at time of diagnosis > 10 ng/ml.
- Prior treatment for prostate malignancy.
- Showing evidence of advanced PCa.

Data-collection methods

Data collection took place from April to November 2013 in private rooms within hospital trust premises or men's homes. Each participant was interviewed once, and LR conducted all bar one interview (which was conducted by JW). Men's partners were present in four interviews (see *Table 17* for details). Where present, partners were considered as participants and subjected to the same consent processes as the men. Interviews were recorded following receipt of written consent from all present and transcribed in full using standard notation. Interview times ranged from 20 minutes to 1 hour and 5 minutes for clinicians, and 18 minutes to 1 hour and 36 minutes for men. Interview schedules were used to ensure topics were consistently covered (see *Appendix 3*). These documents evolved as data collection proceeded, with inclusion of new topics and omission of prompts that were known to be unsuitable/irrelevant based on initial interviews conducted. Clinician/nurse interview schedules explored the criteria informing decisions to offer AS, current AS protocols and perspectives on the AMS. Interview schedules for men on AS covered pathways leading to joining an AS programme, experiences of being monitored and perspectives on the AMS. Brief field notes were taken by the researcher during and after interviews, which recorded pertinent details to pursue within interviews and initial analytical thoughts (if time permitted).

Vignettes displaying three hypothetical men's PSA values over time were used throughout the interviews with clinicians/nurses and men on AS. Each hypothetical man's data were presented in two formats: as a table on paper showing PSA values over time ('paper vignettes'), and as they would appear as a graph within the AMS, presented on a laptop using Microsoft Excel® 2013 software [(Microsoft Corporation, Redmond, WA, USA); *Appendix 3*, 'Active monitoring system vignettes']. Each of the three hypothetical vignettes presented distinct PSA trends: vignette 1, where the values were steadily increasing ('Mr Smith'); vignette 2, where the PSA values were erratic ('Mr Jones'); and vignette 3, where there was a gradual rise in PSA with a steeper gradient for the last few readings ('Mr Evans'). This final vignette was the only scenario in which a man's PSA eventually exceeded a value of 10 ug/ml.

Presentation of the AMS using Excel allowed for addition and removal of various reference lines, each of which were individually superimposed on PSA graphs by the researcher (with a brief explanation of what each reference line represented). Addition and removal of these reference lines allowed the researcher to explore different presentations of PSA values within the AMS.

A core component of the qualitative work was to investigate current AS provision from the perspectives of men and clinicians; this was a broad aim, with scope for exploration of new lines of enquiry. Given this, data saturation was assessed primarily through consideration of the principal aim of the qualitative work in the context of the wider project: to evaluate the AMS tool.

Analysis

Audio-recordings of interviews were transcribed in full using standard notation. Transcripts were analysed thematically using the constant comparison method derived from grounded theory methodology.¹⁰⁷ This involved line-by-line coding of transcripts, categorising codes into themes and developing codes and themes as transcripts were reread in light of newly collected data. Analysis was primarily conducted by LR using NVIVO (V.9) (QSR International, Warrington, UK). A sample (10%) of transcripts from clinician and patient interviews were independently analysed by JW midway through data collection. Any differences in coding and thematic interpretations were discussed and additional areas were suggested for addition to the topic guides. Descriptive accounts of observation and interview findings were written and discussed with the research team. Matrices for major themes from interviews were drawn up. Participants were arranged according to centre and – in the case of clinical professionals – according to role to enable patterns of meaning within each group of participants' accounts to emerge, facilitating identification of 'negative' cases that conflicted with emerging theories. These exceptions were revisited by rereading transcripts and were described accordingly in reported findings. Final matrices of the main themes described in this report were constructed prior to writing to ensure that the full range of perspectives was reported.

Results

Participant characteristics

A total of 38 interviews were conducted (13 consultant urologists, two specialist registrars in urology, three urology nurse specialists, 20 men). All clinicians who were invited to take part in the research agreed, and three of the 23 invited men declined (without giving a reason). None of the participants withdrew from the research. Characteristics of men who participated can be found in *Table 17*. The sample of men interviewed was spread across the age range of 55–70 years, with most aged 65 years and above (as is typical in routine practice).

Presentation of data

Quotations from interviews have been selected on the basis of how clearly and succinctly they illustrate the dominant themes to emerge from the research. Tensions and inconsistencies have been discussed, and quotations from divergent cases are presented where relevant.

Clinicians are referred to as 'he' throughout, for purposes of anonymity.

The following identifiers have been used for quotations:

U(number) = consultant urologist or specialist registrar in urology; SPN(number) = specialist urology nurse; P(number) = man on AS; P(number)partner = partner of man on AS; INT = the interviewer (LR/JW)

TABLE 17 Characteristics of men interviewed

Participant identifier for men	Age at time of interview (years)	Time on AS (years)
P1	66–70	≤2
P2	66–70	≥5
P3	61–65	≤2
P4	66–70	≤2
P5	66–70	2–5
P6	≤60	2–5
P7 ^a	61–65	2–5
P8	66–70	2–5
P9	61–65	2–5
P10 ^a	66–70	2–5
P11 ^a	61–65	≤2
P12	66–70	≤2
P13	≤60	2–5
P14	61–65	≤2
P15	66–70	≤2
P16	66–70	2–5
P17	61–65	2–5
P18 ^a	66–70	≥5
P19	66–70	≤2
P20	61–65	2–5

a Denotes partner present.

Overview

The chapter covers two broad areas explored within this research:

1. deciding whether or not to follow an AS programme
2. current practice within AS programmes: follow-up procedures and triggers for change of management.

For clarity, data from interviews with clinicians/nurses and men on AS have been presented separately within each section. The term 'AS' rather than 'AM' was used by all sites, although clinicians did not make any distinctions between the two terms. They did however distinguish 'WW' from 'AS', with the main difference being that WW did not have a curative intent. A more detailed description of how clinicians distinguished AS from WW can be seen in *Appendix 3*.

Deciding whether or not to follow an active surveillance programme

Clinicians' views on candidates for active surveillance

The early stages of clinician interviews focused on decision-making surrounding in what circumstances it is appropriate to offer AS to a newly diagnosed man. Clinicians tended to weigh up a range of clinical factors in their decision-making, including tumour Gleason score, the volume of cancer detected and PSA values. However, a lack of evidence made it difficult to be definitive about who was eligible for AS on the basis of the clinical parameters currently considered. Existing clinical guidelines varied, and long-term outcome data were not yet available:

I mean I have to say guidance varies depending on what you're . . . or should we say evidence varies, and if you read one paper it says anybody less than five [biopsy] cores and if you read another paper it might say sort of three cores, so you're sort of going for a low core count of about two or three cores.

SPN2

There's a lot of it that's a bit made up, you know, best guess, rather than evidence.

U9

In light of the above uncertainties, clinicians found it difficult to be specific about the upper limits of clinical parameters that made a patient definitely 'eligible' for AS, with the majority indicating that setting fixed thresholds was not reflective of their routine practices. Clinicians were much more inclined to discuss values they were comfortable with and values for which they would be hesitant to offer AS, but made it clear that these reflected their general practices rather than hard and fast rules. Comparison of clinicians' accounts in relation to each clinical parameter (e.g. Gleason score, number of positive cores for cancer) suggested variation across and within sites. *Appendix 3* sets out a more detailed description of clinicians' reported levels of comfort surrounding each clinical parameter, whereas *Table 18* summarises clinicians' preferred maximum values. In some cases, clinicians' thresholds varied within sites, in which case the range of responses for that particular site has been reported.

It was apparent that variation was more pronounced for some clinical parameters than others. For instance, most clinicians were comfortable offering AS to Gleason score of 6 patients, and most expressed willingness to include some Gleason score of 7 patients as long as these were no higher than Gleason score of 3 + 4. Although a minority, some clinicians (from multiple sites) expressed concern about the suitability of Gleason score of 3 + 4 patients for AS:

Some patients with a Gleason 7 low volume low PSA do continue on active surveillance, but it always makes me nervous when it says Gleason 7.

U5

TABLE 18 Clinicians' reported preferred maximum thresholds for clinical parameters considered in decision-making surrounding AS eligibility

Clinical parameter (preferred maximum value)	Site 1 responses	Site 2 responses	Site 3 responses	Site 4 responses	Overall range of responses
Gleason score	6–7 (3 + 4)	6–7 (4 + 3)	7 (3 + 4)	7 (3 + 4)	6–7 (4 + 3)
Volume of cancer: number of positive cores (out of 12)	4 to 'does not matter'	2 to 6	2 to 6	2 to 3	2 to 'does not matter'
Volume of cancer: percentage of core involvement (%)	50 to 'does not matter'	40 to 50	20 to 40	20	20 to 'does not matter'
PSA value (ng/ml)	10 to 20	8 to 'does not matter'	10	10	8 to 'does not matter'

At the other extreme, although Gleason score of 3 + 4 was considered the upper threshold for most clinicians, one urologist also showed willingness to consider Gleason score of 4 + 3 patients – considered by most to be too high a risk group for AS:

I mean obviously 3 + 4 you're taking less of a risk. Um but I think it's reasonable to do a 4 + 3.

U10

Although views surrounding Gleason score were largely similar across sites and individuals, specified figures for acceptable biopsy cores positive for cancer were highly variable, ranging from '2/12' cores to '50% of the cores'. Determining a clear upper threshold for the number of positive cores was not straightforward, as this figure could be compensated by lower percentages of cancer present in each core, or lower values in other clinical categories (e.g. Gleason score):

Essentially if we were going to tie it down we would say no more than 3 cores, no core more than 50% Gleason 3 + 3 and then we broaden a bit by including some 3 + 4s, occasionally might let somebody in, there was tiny amounts in 4 cores for example, we might take somebody who had a single core on the out of 10 that had 60% . . .

U1

Overall, clinician interviews revealed that decision-making regarding AS eligibility in practice was a process of weighing up multiple clinical parameters. Importantly, however, decisions of whether or not to start an AS programme were overlaid by consideration of factors other than cancer characteristics, including: co-morbidities, patient age, patient preference and psychological status:

[. . .] It's an art, this, it's not a science.

U10

Decision-making through triangulation of numerous factors could be viewed as a response to the uncertainties in the field, particularly given that all the clinicians interviewed discussed limitations surrounding how accurately individual clinical parameters characterised PCa; PSA testing, DREs and biopsy were all fraught with limitations:

I think at the end of the day it's the combination of all of them really. We know PSA is not fail-safe, we know DRE's not fail-safe, so at the end of the day a lot of active surveillance is your clinical judgement of putting all the pieces together, so we haven't got a perfect test but we've got a number of things that we can look at, put them all together to give us as much information as we can.

U4

In light of the uncertainties surrounding methods of measuring PCa status, clinicians repeatedly discussed the necessity of weighing up all such factors in adopting a multifaceted approach to management.

Men's decisions to opt for active surveillance

Men were asked to describe how the decision to opt for an AS programme had come about, once they had discussed their routes to receiving their diagnosis of PCa (see *Appendix 3*). The extent to which men took ownership of the decision to start AS varied. Men were broadly categorised into one of three groups: (1) those who chose AS; (2) those who followed the clinician's decision; and (3) those who agreed to AS. It should be borne in mind that this crude categorisation may have been influenced by the clinician's approach to presenting information, individual differences in men's preferences for shared decision-making, as well as by the potential biases associated with men's recollections of events. Despite this, a broad pattern of decision-making emerged for each site in terms of the extent of patient involvement in initial treatment decision-making (*Table 19*).

In site 3, men generally placed themselves as the main driver behind the decision to opt for AS and recalled taking time to research and think about their options. Some men from site 3 spontaneously commented that they had felt their clinicians actively refrained from recommending a particular route for managing their PCa:

... they said 'We cannot recommend anything.' From what I could gather it was more than their job was worth, to recommend anything to me. 'Cos if they get it wrong I could come back and sue. So it was literally left down to myself [...]. It's either this, this, or this. You are left very much to searching out yourself, and making your own decisions. So I went through everything.

P6

Although some men from other sites also reported 'choosing' AS themselves ($n = 2$), they also recalled receiving treatment recommendations from their clinicians. One participant still framed the final decision as the result of his own decision-making, aided by research and discussion with his partner. The other participant (P2, below) opted for AS, but had received a strong recommendation for surgery. This was the only man who reported going against clinician guidance:

And I think originally he wanted me – he suggested that because of my age that I actually have the operation. Probably the best outcome for me, perhaps, would have been to have it removed; being that it was contained in the prostate. But I actually thought about it and did a bit of research on it and I decided for me, I wanted to not have the operation and do the active surveillance.

P2

TABLE 19 Decision-making leading to initiation of AS programme

Site	Men who chose AS	Men who followed clinician's decision	Men who agreed to AS	Total
Site 1	1	3	1	5
Site 2	1	3	4	8
Site 3	6	0	1	7
Total	8	6	6	20

Men from sites 1 and 2 were more likely to distance themselves from the treatment decision-making process, framing the decision to opt for AS as following the clinician's recommendation, or accepting the clinician's decision. There were subtle differences in these two categories. Men who reported having agreed to a clinician's recommendation of AS still expressed the final decision-making responsibility as being their own, or at least acknowledged that they had a role in decision-making. These accounts emerged from both sites 1 and 2:

Just he said his recommendation was I went on watch. It wasn't a problem [. . .]. And I didn't have any problems with that.

P20

In contrast, the final subgroup of men did not allude to there being a 'decision-making process', instead simply accepting the clinician's decision that they would begin AS ($n = 6$). Most of these men were from site 2 ($n = 4$).

Although the sample numbers were limited, these data indicated possible variations in practice across sites and clinicians when it came to shared or informed decision-making in the context of treating low-risk PCA. However, factors such as sociodemographic differences across patient populations and patient recollections/response styles must be borne in mind.

Current practice within active surveillance programmes: follow-up procedures and triggers for radical treatment

This section will consider AS protocols from the perspective of clinicians and men. It should be noted that sites varied in how protocol-driven their practices were. At one site (site 1), clinicians had been following a site-specific protocol for a number of years, whereas clinicians from another site (site 2) were in the process of developing a protocol at the time of data collection. As such, there was more variation across clinicians' accounts from site 2, although it became apparent that individual variation between clinicians existed irrespective of the presence of a protocol.

Clinicians' reported follow-up practices

Follow-up routines for monitoring patients on AS consisted of regular PSA testing, DREs, rebiopsy and face-to-face discussion with the patient about any new or developing symptoms. Detailed accounts of the frequencies of each of these measures can be found in *Appendix 3*. Overall, the frequency of tests and measures varied across sites and, in some cases, among clinicians from single sites. All clinicians highlighted that published guidelines for 'optimal' follow-up routines varied, which in turn reflected the lack of long-term evidence to inform ideal AS provision:

I mean I think [. . .] I think what I would admit to is saying that [. . .] we [. . .] there often isn't a consensus [. . .] as to how these people should really be seen. And I think that's a consensus not only nationally, of course, but [. . .] also within a, within a department. So I'm sure if you were to actually look at the surveillance [. . .] management in this department it may well be different according to which person you'd see.

U9

Despite the lack of conclusive evidence, sites 1 and 3 had a site protocol in place to lend some structure to their follow-up practices. Most clinicians from site 1 quoted this protocol while discussing follow-up routines, although some clinicians were explicit about deviating from this protocol at times:

So there's the PSA follow-up and then we do rectal examinations (6 monthly) – well personally I don't do rectal examination in every clinic, I would try and do at least one rectal examination per year in a patient.

U2

The idea that site protocols were to be viewed as 'guidance' that permitted deviation was also apparent at site 3. Only two clinicians were interviewed from this site, one of whom acknowledged that the protocol was not tightly enforced. The second clinician's interview also emphasised that follow-up practices were subject to individual clinical opinion and/or practice:

INT: So in terms of the actual routine of active surveillance, what does that entail, then, in [Site 3]?

U15: It varies from consultant to consultant, but for me it'll be regular . . .

Reported frequencies of certain clinical tests/measures in AS follow-up varied to a greater extent than others. For instance, the frequency of PSA testing (via general practice) and frequency of DREs (via outpatient appointments) could be considered as one of the more fixed components of the AS programme for most sites, with some examples of individual variation in reported practices. Most variation for these clinical measures was reported from site 2, which was in the process of formulating a centre protocol. Reported practices in site 2 suggested indefinite intervals in the frequency of PSA tests and polarised opinions regarding the value and role of conducting DREs in follow-up:

U8: So following an outpatient discussion on the available evidence that that's what we're going to do, then the patients will have another PSA 3 months after that. And then depending on the results, there will be a review of the PSA 3 to 6 months after that.

INT: I see, I see. So in terms of DREs, how often do you do those?

U10: We wouldn't. . .

INT: OK . . . and do you think there is any value in doing that?

U10: No I don't, no, not at all.

INT: That would explain why you're not doing them.

U10: No I mean I guess if you had someone who had got a palpable abnormality you'd probably want to keep a closer eye on that, but for the majority of patients, no [. . .]. But if you're doing a biopsy every 2 years I don't think there's a great value in doing a DRE regularly. [It's] not evidence based. It is evidence based for follow-up to [active] treatment, post-radical treatment – but not for surveillance.

U9: And, you know, there is this big difference: T1C disease, impalpable disease is clearly easy to monitor through PSA, but if they develop a T2 prostate something is changing and, you know, that's – it's easy to diagnose that and you should be (aware of it). It's like, I don't know, treating blood pressure without actually measuring the blood pressure.

In contrast to PSA testing and DREs – both of which were relatively non-invasive – the frequency of rebiopsy appeared to be a more controversial topic across clinician interviews, prompting the greatest expression of uncertainty. Regardless of whether or not a site protocol existed, all clinicians conveyed the delicate balancing act that was required when it came to decision-making surrounding rebiopsies, given the risks of comorbidity and patient anxiety:

One it's unpleasant and two if there's morbidity attached to it, there is a mortality attached to biopsies, 0.1% or something, we've had one patient who died about 3 years ago and this year, over the last 12 months, we've probably had about four or five who've been admitted with sepsis [. . .]. That's why I think it's very difficult to say 'you've got to have a biopsy every year or every 2 years' but I

think you have that as your basic requirement, which you can then alter a bit depending on the patient and the disease dynamics.

U2

Sites with established protocols once again yielded the most consistent reported practices with regards to rebiopsy. However, clinicians' tendencies to view site protocols as guidelines were reiterated here, with accommodation of patient choice and comorbidities perceived as being important:

Some men for whatever reason have a biopsy earlier than a year perhaps because either the patient or we are a bit less happy about the approach maybe and some men, a minority, decide they don't want a repeat biopsy.

U3

[. . .] Biopsy regimen is usually 18 months or 2 years. And I tend to do it at 18 months or 2 years depending on the man and their health and what they're like.

U15

Similar to other elements of AS protocols, the greatest variation in rebiopsy practices came from site 2, where clinicians were explicit about the need for a protocol. Variation was found in relation to 'initial' rebiopsy times (ranging from 6-monthly to yearly) and subsequent rebiopsies (ranging from yearly to 'based on need'):

Yeah – er, well we will rebiopsy everybody, even if the PSA is stable [. . .]. So if there is no change in the grade or the number of cores or indeed the percentage of those cores involved, then we will continue with PSA surveillance. What we haven't tied down is what to do about the next biopsy.

U8

Well we would, for a typical man on surveillance we would generally rebiopsy them within 6 months. And then after that they would generally continue with 3-monthly PSA monitoring, and we'd see them fairly regularly to repeat (their rectal) exam. And then ongoing it's a bit more difficult: there aren't any hard and fast rules. But again the only reliable way of detecting tumour change is to do a biopsy, so I think you are looking at doing a repeat biopsy every couple of years [. . .]. But we don't have a hard protocol, because nobody can agree.

U10

Clinicians' thresholds for moving off active surveillance

The final element of AS protocols that was essential to understanding current practices was the trigger for moving from AS towards radical treatment. The uncertainties underpinning routine practice for AS follow-up also emerged here, where there was a lack of evidence to inform decision-making about radical treatments. Clinicians were explicit about this uncertainty, although all were able to discuss triggers that alerted them to consider radical treatment:

The PSA doubling time 3-year cut-point has just been proposed by [named eminent clinician] because he did use 2 years and found that a lot of patients were doing very badly. So he threw it out to 3 years, just as a bit of guesswork, but that seems to have stuck with practising urologists.

U11

The single clinical measure with consensus for a clinician's recommendation to consider radical treatment was disease progression as indicated from rebiopsy results. This was reported in clinicians' accounts of routine practices but also emerged in their responses to vignettes. None of the clinicians was prepared to recommend radical treatment to a patient on the basis of PSA measures alone, regardless of whether or not they were presented in the paper-based vignettes or in demonstration of the AMS. Where clinicians were concerned about PSA trends or values presented in the vignettes, their proposal was always to

rebiopsy the patient. Biopsies were perceived to be the most informative investigative tool currently at their disposal:

The ultimate decider other than the patient's choice, the ultimate decider is, is the biopsy.

U1

Clinicians' reliance on rebiopsy findings in decision-making did not signify a widespread enthusiasm for this procedure. Biopsies appeared to be the best possible option given the lack of alternatives. Clinicians unanimously expressed a lack of confidence in PSA behaviours as a proxy for cancer progression, citing the many variables that could influence readings, controversy in existing literature, and personal experiences where PSA values had failed to signal serious disease progression:

I think I'm very pro-rebiopsy for men who are on surveillance, because PSA is not – we know that PSA alone is not reliable. Because we know there are so many factors that influence a PSA, and it might be that chap has done vigorous exercise for a day or so before, he's had sexual activity or, you know, any of those things. So, for me, a rebiopsy is essential, yeah.

SPN3

I mean, there is this thing, there are at least two studies that throw quite considerable concern on PSA surveillance, the reliability.

U11

There were many factors that prompted clinicians to offer an unscheduled rebiopsy to patients on AS, including changes in DRE results and/or patients' reported symptoms. Generally, however, most clinicians referred to 'concerning' PSA behaviour as a prime trigger, although there was once again uncertainty surrounding what this constituted. In general, clinicians tended to look for trends in PSA activity rather than absolute figures to trigger rebiopsy. In one exception, a specialist nurse suggested that there was significance to a PSA reading of 10 ng/ml, but explained that his views had been influenced through knowledge of the ProtecT trial protocol (which included additional investigations if PSA was ≥ 10 ng/ml).¹⁴ The only other clinician (U9) to mention the significance of absolute PSA values also saw 10 ng/ml as a significant threshold, but went on to explain that this was merely a figure they were conscious of given its publicity. PSA trends were thought to be of greater value:

U9: I explain the fact that the PSA is a biological variable and therefore it's going to go up and it's going to go down and, you know, we're looking for trends rather than absolute numbers. And I move them away from that – you know, the fact that if you hit 8 or if you hit – you know, it's not an absolute number.

INT: OK that's what my next question was going to be: whether there was any particular . . .

U9: In my mind there are absolute numbers, but I would never tell the patient that.

INT: I see, I see. In your mind, what absolute numbers?

U9: Well there's a lot of talk about the PSA over 10. But of course it depends where it's come from, and so the rate of change is more important.

Further exploration of what constituted a significant trend in PSA rise revealed a range of responses, although all but one clinician framed this in terms of PSADT. Reported thresholds for 'concerning' doubling time ranged from 2 years to 5 years across interviews. Despite the presence of protocols at some sites, individual urologists' thresholds for doubling time still varied. For instance, all clinicians below were interviewed from a single site, which specified doubling time in its protocol for considering radical treatment:

U2: In theory it's the doubling time in 2 years I would say.

U3: So PSA doubling time is the main thing that we use here, so PSA doubling time of less than 3 years would be broadly speaking be the kind of cut off for us.

U5: Well anything under 5 years is considered significant.

A number of clinicians were vocal about the uncertainties surrounding the thresholds they used, explaining that these were fairly arbitrary:

Also we calculate PSA doubling time and again, nobody knows what it means.

U1

There were also variations in how clinicians arrived at doubling time values. Some calculated this mentally, some estimated by looking at the gradients of graphs, whereas others were able to generate values using computer software (available on site 2 systems only, but accessible via the internet by clinicians at other sites):

Yeah, we don't work it out . . . I mean, you eyeball it.

U6

Yeah, so you need it to be done for you and the result software that we have does that.

U3

Although less frequently reported, a number of clinicians talked about comparing actual PSA readings with expected readings, based on the assumption that PSA values would theoretically increase by 5–15% a year as a result of benign prostate hyperplasia (an approach based on similar principles to the reference ranges for men without PCa shown within the AMS). PSA figures that deviated from this extrapolation were considered to be worth investigating if confirmed on repeating the PSA. These clinicians emphasised that they would not act on one PSA reading alone:

[. . .] forgetting the cancer, the benign component will give you a rise in your PSA of something approaching 10% per year. So if they're broadly in that line. . . [. . .] . . . well a jump in the PSA, whatever that means, would alert you to do something different, but the first thing would probably (be) to let a little more time go by, maybe another month and do another PSA on the proviso that they've got no other symptoms.

U1

Finally, clinicians reported that patient choice was one of the main triggers for moving men from AS to treatment. This was thought to account for a substantial proportion of cases where men had moved off of AS, and was reported by most clinicians at some point during the interview:

One in five of them will have to move because their disease will change, but also that probably up to another one in five will move because they become unhappy with the concept of Active Surveillance, so roughly half of the people who drop out of Active Surveillance, drop out from good pathological reasons; half drop out because psychologically they haven't been well enough prepared and people in

the family have a go at them and you know you tell somebody they're on fire, they want a fire extinguisher to put it out, tell somebody they've got cancer and they want an operation.

U1

Patient choice and empowerment are notions that are promoted across the NHS. However, there was a sense that these were particularly important in running a successful AS programme. Clinicians stressed that one of the key goals of AS was to improve quality of life. Anxiety over PSA testing or disease progression could counter the benefits of avoiding radical treatment and its side effects; as such, patient anxiety or distress was a key trigger for considering alternatives to AS:

And, you know, if you're going to make this decision to have no treatment, you've got to be comfortable with it or we have actually created a problem, not solved a problem. It is all about their quality of life and their length of life, and relative between the two which one matters most.

U9

Clinicians' accounts suggested that patients' choices were central to shaping treatment decisions in their practices, but this was also regularly reflected in clinicians' responses to vignettes, particularly where a site protocol was not available (site 2). Accordingly, evidence of a pattern emerged at site 2, whereby responses to hypothetical scenarios were often non-specific and shaped around patient choice foremost:

All these things, there are no cut-offs. All a continuum and, you know, we have the parameters that we use to help make a decision, and then the patient will – will decide.

U8 (in response to a vignette)

By contrast, although clinicians from site 1 (where there was strong adherence to a protocol) also discussed patient choice as a key decision-making factor, their responses to vignettes were often more concrete and definitive than those of clinicians from site 2 and were framed around the clinician's opinion foremost:

I mean if he'd been biopsied at a year and then 2 years, we'd probably have three or four lots of biopsies now and if they all showed Gleason 6 low volume then even with this rise in his PSA – if he's happy – I would continue monitoring with active surveillance.

U4 (in response to a vignette)

The idea that clinicians at sites without an AS protocol placed greater reliance on patient choice is a hypothesis that warrants further exploration. It was certainly apparent that notions of patient preference and involvement were more emphasised at site 2, where clinicians were more explicit about the uncertainties surrounding their AS programme.

Men's views and experiences of current active surveillance practices

Interviews presented an opportunity to build an understanding of how the experience of undergoing AS impacted men's lives. Men's narratives overwhelmingly indicated the relative ease by which they continued with their day-to-day lives, undisturbed by any practical, physical or psychological implications of their cancer diagnosis:

Um [pause] I'm (yeah just?) learning to live with it. I don't even think about it.

P13

It should be borne in mind that the sample of participants interviewed were not only 'low-risk' PCa cases, but may also have represented men for whom AS had been a relatively smooth experience. However, men's reported low anxiety appeared to be a measured response to how they understood their particular form of cancer and their confidence in the care at their disposal. *Appendix 3* sets out a more detailed account of men's perceptions of having 'low-risk', 'non-aggressive' cancer, which in turn stemmed from

their interpretation of clinicians' verbal and non-verbal cues when discussing their disease status. Further to this, men drew great reassurance from the process of follow-up in AS. Men's apparent relaxed attitude prevailed despite awareness that their cancer could progress and that AS was not a permanent management plan. Nine men associated their 'relaxed' state of mind with their knowledge that they were 'in the system' and 'cared for':

*So, you know, if something was to go, you know, badly wrong, then, I'm in a good place, really.
[Slight laugh]*

P3

Most of the men who discussed this theme had limited, if any, experiences of PSA values that warranted re-evaluation of their treatment plan. However, the theme was also apparent in one participant's account of a period where his AS plan was under question:

Um, all right, I suppose when the thing's, the PSA, started to go up, moving up to 12, whatever it was, they were talking about it. But even [...] that really didn't concern me. Because if you like, I was in the system. It was being observed.

P15

Men's awareness of uncertainty in follow-up protocols

The uncertainties inherent in clinicians' descriptions of follow-up protocols in AS were largely absent in men's interviews. There was a general tendency for men to look towards their clinicians in matters relating to the frequency and nature of their follow-up, placing complete trust in those overseeing their care:

My, my attitude is [clinician] is the man who knows. If he's happy with what he sees and he's happy to keep carrying on the way we are with the active monitoring, I'm happy as Larry.

P19

There was little evidence to suggest that men had experienced or been influenced by uncertain follow-up protocols or inconsistencies in clinicians' practices. This may have been a consequence of some men consistently seeing the same clinician or of the possibility that clinical characteristics of the patients interviewed had remained in a zone in which clinical opinion was consistent and/or PSA values were relatively stable. There was one exception, where one man and his partner had recognised inconsistencies in clinical opinion with regards to discussion about the usefulness of DREs and frequencies of biopsies:

P10: They're all, I'm sure, very experienced nice people. But they all seem to tell a slightly different thing [...]. There was one man who said the DRE was as important as the biopsy. And he said that's what would trigger me into action. So they do seem to have slightly different ideas.

P10 partner: One thing that they can't seem to make their mind up about is the frequency of biopsies. [...]

The above case was the only example of a man's awareness of the uncertainties overlying his care. On the whole, men appeared relaxed and confident in the care and follow-up they were receiving.

Men's understandings of follow-up protocols

A key theme to emerge from men's accounts of being followed up in AS was their expressed understanding of the need for multiple tools to inform clinical follow-up. This was most evident in the widespread understanding of limitations in PSA testing and the subsequent need for other follow-up tools (e.g. DRE and rebiopsy). All but one participant showed awareness of the limitations of the PSA test. Three of the dominant themes to emerge included the idea that (1) PSA testing is an indicator rather than a

direct measure of disease; (2) variation in PSA values may relate to causes other than cancer progression; and (3) there was controversy surrounding the meaning and interpretation of PSA measures:

INT: How do you view this PSA figure, then? What do you see its function being?

P1: As one of a number of possible indicators. It seems to be, it's the one that seems to me to be most easily measured.

P9: The nurse who does my PSA tests, her husband had a huge um PSA score but didn't have any problems at all. No, no sign of cancer.

P2: I've heard in the past that doctors don't want to do the test. Because it's a vague test and you might get a false negative, you might get a false positive.

P6: And it is variable. I mean in this country anything above two is considered going up. In America they don't start even considering anything until it's four or five.

Men from all sites had a clear understanding that rises in PSA were not necessarily indicative of cancer progression and were aware of multiple non-malignant factors that could affect PSA readings:

I've accepted it because actually, it's a guide and it's a guide only in terms of what may or may not be happening. And I suppose that idea is reinforced by the fact that I know to a certain extent I have been able to control my PSA readings by doing or not doing something. And they're also saying prostatitis muddies the waters on your PSA. We don't know the impact of it. So for me it's like, I'm just going this is just a broad guide that is applied to my situation. And for me, there is a personal challenge in seeing how low I can get it.

P15

In light of the known limitations of PSA testing, men generally welcomed the addition of DREs and regular rebiopsies (although there were some exceptions, as reported below).

Men's acceptance of follow-up practices

All men participating in this study regularly saw urologists in outpatient appointments. Although they generally valued these interactions, some gave a lay cost–benefit analysis of the practicalities of attending outpatient appointments versus the content and duration of the interaction with the clinician. Men's unprompted descriptions of their experiences of outpatient appointments were strikingly similar, portraying consultations as a brief chat about PSA trends, occasional DREs, with some exploration of how the man was feeling:

No they, they just glance at my PSA and say how are you feeling, and I say I'm fine. Any problems? No. [. . .] Thank you very much. Goodbye. It's probably as quick as that. I can be in and out.

P6

Generally, men were not able to be precise about how frequently DREs took place. Based on their experiences alone, DREs appeared to be spontaneous events. This may have either been a reflection of differences in individual clinicians' practices, or deviation from AS protocols (if these existed):

Yep. Sometimes they say oh, [. . .] pop up and they'll just [. . .] check you out.

P10

A number of men who had experienced brief interactions with clinicians in outpatient appointments (with no DRE) questioned the value of face-to-face consultations, in light of the resource and practical considerations for the patient and the NHS:

P9: There's a school of thought that says do they really need to do that, because at 6 months all they're looking at, really, is the PSA levels and asking me how I am. You're just wondering whether these are necessary, you know, to bother a busy consultant.

P1: If we were OAPs without resources, being able to drive, or be driven, and can afford to go in. Pay the car parking . . . then those meetings, you would really question the value of going in, while we were on active surveillance, if it could be done on Skype or on a phone.

P2: And I find it a bit of a waste of time really. Because . . .

INT: Right. The digital rectal?

P2: No. No. The fact that they don't do that every time. I mean I go out there, I know what my PSA is because I phone the doctor up and find out. I know if it's gone up [. . .] and if they're not going to do an examination, we're just going to talk about the PSA, then I can work that out for myself. Or if they decided from their point of view that they wanted to speak to me, then perhaps they could send me an appointment.

The issue of rebiopsy in AS follow-up elicited strong views among some of the men interviewed (often simply referred to as 'biopsy' during interviews). For the purposes of this discussion, all references to 'biopsy/rebiopsy' refer to transrectal ultrasound-guided biopsies that took place as part of the AS programme.

Men's understandings of rebiopsy were based on a combination of their responses to vignettes and reflections of their own experiences. In terms of their own experiences, most described the procedure and its after-effects as unpleasant, with descriptions ranging from discomfort to extreme pain:

No. All right, not the best of things. A bit uncomfortable, but could be worse. [Laughs]

P2

Oh the sheer pain. And the horror and indignity of it. I don't know whether it was him, but he was just so rough. It was just [. . .] un [. . .] believable.

P16

Despite this, men generally accepted rebiopsies, perceiving these as an essential component of AS. This widespread view went hand in hand with awareness of the limitations of relying on PSA measures alone and the view that rebiopsy was the best available indicator of cancer progression. Similar to clinicians, patients' responses to vignettes often culminated in the conclusion that a rebiopsy would be required:

Yeah. I mean, I think the PSA is not to me, a fixed, sort of, super accurate measurement. So I think, the biopsies are much more important as far as I'm concerned.

P13 (in relation to paper presentation of Mr Smith vignette)

Although men described rebiopsy as the most accurate indicator of cancer progression, rebiopsy was also the single intervention that some men had declined. Four of 20 men reported refusing or refraining from at least one rebiopsy that was scheduled as part of their AS plan, either because they felt the benefit did not outweigh the risks of side effects or because they fundamentally disagreed with the protocol-driven nature of such an invasive intervention. Some of these men had experienced side effects from earlier biopsies:

And I said, you know, I'm not going this route again. OK, biopsies, like the way I've had them taken, is not gonna be an option. If I'm going to get this sort of impact from it, it's not worth the pain and – because I don't get any pain from the cancer, I don't wanna go there [. . .]. Why have a procedure that actually causes you more grief and pain than er . . .

P15

Two men disagreed with the protocol-driven nature of rebiopsy, preferring a more targeted approach. In particular, one man felt his PSA readings did not support the need for rebiopsy:

My last consultation was in July. After the 3.9 came back. Which I was pleased with and my doctor was pleased with. And they said 'Right, we want to run some more biopsies.' And I said, 'Well I don't agree,' and I told them specifically no I don't agree with this.

P6

Men who had refused or delayed rebiopsy felt that they would resort to this intervention if their PSA values started to cause concern. This was also reflected in responses to vignettes:

I mean his PSA tests, well, with taking a look at that [mid increasing trend] sort of level I would maybe begin looking at maybe another biopsy.

P6 (in relation to paper presentation of Mr Evans vignette)

All four men who had refused/delayed rebiopsy had expressed knowledge that PSA was not a fail-safe measure of PCa extent or progression. Paradoxically, however, relatively steady PSA values appeared to offer sufficient reassurance to refrain from rebiopsying:

I know it's my decision not to have it at the moment. But I've had to try and balance the odds and look at the statistics and say while my PSA's reasonably stable and all that sort of business, then that's fine. If my PSA had gone up [. . .] even that [. . .] it's a significant but a small amount, I would've just said 'Right, OK, I now understand I've definitely got to have it.'

P9

It could be that I was also extremely lucky and I actually have terrible cancer and they only found one, tiny core. But then the PSA level would reflect [. . .] that fact. So I think it's taking a look at both things, and using a bit of [. . .] common sense, if one dare say.

P6

Some men ($n = 7$) talked about magnetic resonance imaging (MRI) scans as an alternative to rebiopsy. Two men had heard of new developments in MRI scanning from peers and the media and intended to ask clinicians about this possibility. The remaining five men were aware that this was not routinely available in their current NHS trust. Only one of these men (P1) had directly asked his clinician about this. P1 had accepted the clinician's response that there was not yet sufficient evidence to support use of MRIs, but also appreciated the clinician's explanation that resource-related issues may prevent access to this technology:

And he was very, very honest. You know, I was able to anticipate with saying, you know, what they're able to do in [a leading research hospital], and what we're able to do in the [current location] may be

different in terms of both, you know, the finances involved, and the equipment, the availability. And one understands that, of course.

P1

Other men assumed that MRI scanning was not routinely available owing to resource implications, although one of these men (P15) had received an MRI scan on one occasion.

Men's triggers for moving off active surveillance

Men's potential triggers for moving from AS to radical treatment were explored through direct questioning and discussion of vignettes. For most men, the decision to move off AS would be directed by the clinician's recommendation:

I asked him, last time, because I've made that decision, does that hang over my head for years? Or do they say you should really be thinking of the next . . .? And he said oh, he said straight away, he said 'oh no. That's my job,' he said. 'I'll tell you when you get to a point when you should be rethinking it.'

P7

P12 (in relation to paper presentation of Mr Evans vignette): I'd still probably want to wait. A bit. Yeah – see if that trend continued, probably for another [. . .]

INT: Would you have a cut-off point?

P12: Yeah, that's, it's an interesting question. Because that's the tricky one, isn't it. That's what the doctors have to always judge. Yeah.

In many cases, men did not engage with the prospect of making these decisions themselves, indicating that they had never given their personal triggers consideration. For example, the participant below framed his response in terms of what the *clinician* may or may not do, despite being asked about his triggers. This style of response was typical of men, reinforcing the idea that they viewed clinicians as the principal driving force behind such decisions:

INT: Yeah. Do you have any inclinations as to what might be a trigger for you, as an indication that it's getting worse? I mean you mentioned a sudden kind of jump up in PSA.

P14: Yes. Well, I'd be guided by them. I mean if they said, ooo yes, your PSA's gone way up, you need to [. . .]. Well I imagine they wouldn't um – they'd do another examination. They wouldn't do anything just based on the PSA level. They'd have to come and have another look at it.

A few men were more specific about triggers that clinicians had previously mentioned, but, even in these cases, the responsibility for such decisions was squarely placed on the clinician:

P17: I think they said they don't [. . .] broadly consider surgery until it hits 10. Something like that. 10 to 12, I think it was they said.

[Later]

INT: You mentioned a PSA of 10. If you did hit that, what do you think you would do?

P17: I mean, again, I'd be looking to be guided by the consultants. I mean that's what they're there for. In my view.

Two other men were distinct from others. Although both suggested that clinician input was important, they had clearly given some thought to their personal thresholds, although these were expressed in very vague terms, without mention of PSA values:

I sort of think my body would kind of let me know if things were getting a bit serious [. . .]. I am one of these super-sensitive body-aware types.

P13

P6 (in relation to paper presentation of Mr Jones vignette): *If you're 55, it's certainly – it's not something that I will consider unless I really, really had to.*

INT: *That was going to be one of my next questions. Under what conditions do you feel you might?*

P6: *If they told me it was life-threatening.*

INT: *If they told you?*

P6: *Yes. Or I thought it, myself, was life-threatening.*

The overarching theme of 'looking to the clinician' in decisions regarding staying on/moving off AS was further reflected in men's descriptions of their own experiences of receiving and interpreting PSA test results. The dominant theme in men's accounts was the focus on how clinicians interpreted changes in their PSA values. The data presented here were all taken from unprompted accounts; although men had not been asked what their clinician thought, 12 participants spontaneously referred to following the clinician's lead in interpreting PSA values:

I mean, because I mean it's . . . it's not worried me because I know that it's pretty steady and I get reassurance – you know, if, when I go in and the guy, the consultant says, well it's pretty steady around about three. We're happy with it. And I say, well I'm happy with it.

P9

Well, I haven't the faintest idea what my PSA results are. And to be honest, at this stage, I'm not really bothered – as long as the monitoring is OK and things are picked up. That's fine by me.

P19

It was clear that patients' impressions of clinicians' attitudes were integral to shaping their overall picture of how well they were doing on AS:

Well yeah, the last high point was when he said that you're 4.2. And he said for a man of your age, now this is a brilliant result.

P5

I'm back exactly where I started. So I, I've done that. My graph goes that way. So they're quite chilled out about me.

P15

They seem happy. So if they're happy I'm happy.

P14

The above patterns were also apparent in men's descriptions of rises in PSA. The men below framed their accounts of experiencing a rise in PSA using recollections of clinicians' explanations:

*And then it came up to 3.5 and then 3.6. But the doctor said, you know, those **little** bits that can be anything can trigger that. Nothing to worry about.*

P3

I suppose it is creeping up. Over 2 years it's probably gone up about [. . .] 3 decimal points, I suppose, something like that. But I'm told that's quite normal.

P9

There were a few examples where men had picked up on subtle, sometimes unspoken, cues from clinicians to piece together an interpretation of their suitability to continue on AS. For example, the participant below interpreted the clinician's body language and physical actions within the consultation as an indication that there was nothing to be concerned about; similarly, the man represented by the quote below this reported what he perceived to be a change in his clinician's attitude in the last appointment he had attended, leading to some anxiety that his cancer had progressed:

You, 'cos you tend to look at them and, you know, and if you [. . .] Think I'd say, you know, I think they're probably quite good at hiding things, but you know, they all seem to be quite, you know. It's just a case of going in, PSA levels look good, you know. And they're writing it up, actually, as they're talking to you. And you think, well, it can't be anything too bad while they're . . .

P9

It's his phraseology was different this time. Umm it's 2.9, now and it has been going up slightly, this last year. Or something like that. And his – he mentioned the word cancer, which he never did before.

P18

Finally, although men generally looked to the clinician to interpret their continued suitability for AS, many were aware of the types of triggers their clinicians looked for. Men's spontaneous accounts of what their clinician may be concerned about have been outlined in *Appendix 3*. Regardless of this knowledge, all of these men still maintained that the final decision to stay on or move off AS would be based on a discussion with their clinician.

Conclusion

A key finding from this study was the considerable uncertainty experienced and expressed by clinicians and men. Uncertainty emerged in relation to the underlying evidence base for optimum practice in AS and clinical decision-making in practice. Key areas of uncertainty included decisions surrounding who should be offered AS, optimal AS follow-up protocols and decisions relating to when it is appropriate to consider moving from AS to radical treatment. This uncertainty was reflected in clinicians' tendencies to refrain from being specific about clinical parameters and the variation in clinical practice that existed within and between institutions and individual clinicians.

Men on AS were well informed in relation to the clinical uncertainties surrounding the value of PSA as a prognostic tool, and most accepted the value of follow-up combining PSA measures, DRE and rebiopsy. In general, men used their clinician's verbal and non-verbal cues to conceptualise their diagnosis and current PCa status and placed a great deal of trust in their clinician when it came to making potential decisions about moving on to radical treatment. However, some men used favourable PSA trends as justification for decisions to refuse certain elements of AS follow-up protocols (rebiopsy), despite understanding that PSA was an imperfect indicator of cancer status.

Clinicians and patients well understood the limitations of PSA as a monitoring tool, and, yet, patients were mostly not particularly anxious or concerned about this. Patients indicated that they trusted the clinicians to advise them appropriately. Clinicians expressed their concerns and anxieties to some degree in the interviews but seemed not to have transmitted these to patients. Future studies should use observation methods and/or audio recordings of routine AS follow-up appointments to consider the content of discussions to investigate this further (see *Chapter 9* for further future research recommendations).

Chapter 7 Qualitative study: clinicians' and men's views on the active monitoring system

Introduction

This chapter will consider clinicians' and patients' views on the AMS, addressing the following objectives of the qualitative study (as reported in *Chapter 6, Introduction*):

- to explore clinicians' and patients' views on how the AMS compares with current practice
- to investigate which presentations of the AMS are most acceptable to clinicians and men.

What follows is a brief introduction to each of the AMS reference lines that feature in the findings. For ease of reading, the reference lines will also be referred to by their colours. The AMS consisted of three reference lines, as shown in *Figure 9*:

- the 'upper reference range for men without PCa' (orange line)
- the 'predicted PSA for a man without PCa' (blue line)
- the 'predicted PSA for a man with PCa' (green line).

The black lines show the PSA results for the relevant vignette [(1) Mr Smith, (2) Mr Jones or (3) Mr Evans]. As discussed in *Chapter 6*, the AMS allowed for addition and removal of the above reference lines, thus allowing the researcher to explore different presentations of PSA data within the system. These different 'presentations' refer to different relationships between reference lines and hypothetical data.

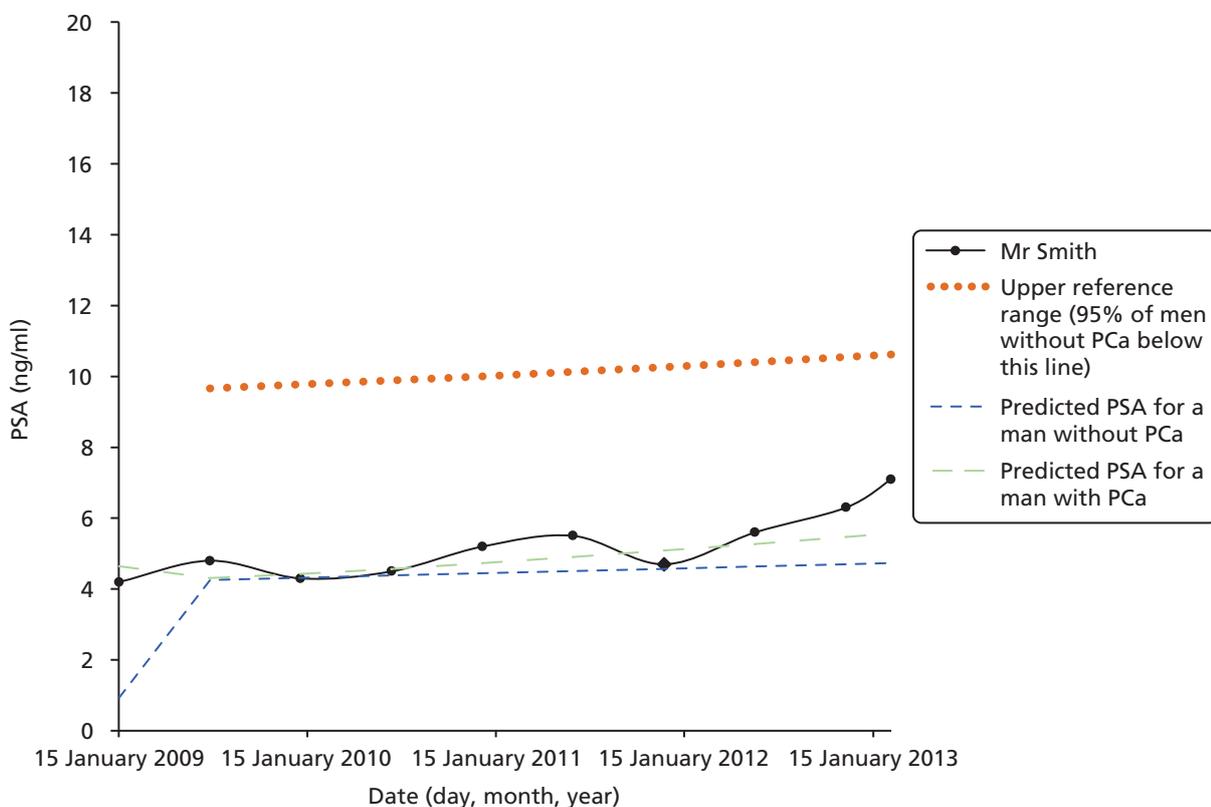


FIGURE 9 Active monitoring system.

Methods

Please refer to *Chapter 6, Methods*.

Results

Details of the sample and presentation of data can be found in *Chapter 6, Results*.

Clinicians' views on data presented in the active monitoring system

Clinicians' views on the upper reference ranges and predicted prostate-specific antigen for men without prostate cancer

Clinicians expressed mixed views about the value of having data relating to men without PCa presented in the AMS. Two out of 18 clinicians/nurses could not see the relevance of these data for a system targeted at men diagnosed with PCa:

But if you know they've got prostate cancer I don't think you need to know what the normal range would be.

U15

Others tended to take a broader view of how the data for men without PCa could be used in practice, particularly in the earlier stages of AS for newly diagnosed men. All of these clinicians viewed the predicted PSA and upper reference range for men without PCa as having the potential to reassure patients but felt that these data were of little relevance to their own decision-making.

Reassuring patients with and without prostate cancer

Two clinicians felt that one advantage of the 'predicted PSA for a man without PCa' (blue) line was its potential to visually demonstrate how PSA naturally rises with age:

U2: I'm just trying to imagine myself in a consultation with a patient.

INT: And would that apply to all of these then?

U2: Yeah, no I think in general for patients who you're trying to explain their PSA. Also the nice thing about this line is it's going up, OK that says a short timescale but again it would help to explain that PSA naturally rises.

Three clinicians suggested that the blue line could be used to reassure men referred by their general practitioners (GPs) for further investigation that they did not need a biopsy. According to two of these clinicians, the fact that the line has a slight gradient, demonstrating age-related increases, was thought to be key to this:

You could say that, because we see a lot of men who've had PSAs over the years and suddenly their PSA has reached a threshold where they get referred, but just 'cos their PSA has reached the threshold, it may be going up on this nice blue line, and actually they don't need another biopsy, they don't need any biopsy at all. Their PSA is just going up on the curve that you'd expect.

U6

Two clinicians felt that the blue line did not present any additional information to what they already used in practice. Similar to the above, these two clinicians also believed that age-related PSA values could reassure patients but commented that they already had access to and used these data in their current practice (although not graphically). The AMS therefore added little in terms of additional information:

U4: In a way we kind of already do this by just looking at the . . . by knowing what's age-related reference ranges for normal men.

INT: Yeah, so you just have that . . .

U4: We kind of have that information in our head, yeah.

Finally, one clinician felt that the 'upper reference range for men without PCa' (orange) line was also a powerful visual aid for reassuring men that they do not have PCa:

I suppose it would be a useful line to illustrate to patients your PSA is still pretty damn low sort of thing and when you're trying to diagnose the disease then it's a bit more useful. So your level's still very low, your biopsies are negative, this is where you are, 90 however many per cent of men have a PSA below that level, which is you won't have cancer, so it fits in, it would help in that respect.

U2

Clinicians' views on showing the predicted prostate-specific antigen for men with prostate cancer

The 'predicted PSA for a man with PCa' (green) line was often the sole focus of the AMS for clinicians who ignored the data for men without PCa, seeing the latter as irrelevant:

The green line would be the key one, because you're talking about men with prostate cancer.

U6

Most clinicians' ideas on how the AMS could be used in practice were voiced with reference to the green line.

Following the line showing predicted prostate-specific antigen for men with prostate cancer is reassuring

Four clinicians suggested that having values in alignment with the line showing predicted prostate-specific antigen for men with prostate cancer (green line) could be reassuring to men on AS. Having a man follow this line was thought to create a sense of being in control, in that it suggested that the cancer is behaving in a predictable way:

U3: The green one is useful because I guess that's kind of what we'd be expecting it to do if it is the disease that we think we've got [. . .].

INT: What do you think would be particularly reassuring about it to the patient?

U3: Well if we're saying that, which you would have been for a number of years, that according to the type of tumour that's been found this is what one would expect then that's really what the green line is giving isn't it.

The influence of the line showing predicted prostate-specific antigen for men with prostate cancer (green line) in clinical decision-making

Four clinicians made specific reference to the green line when discussing implications of the AMS for clinical decision-making, although only one of these clinicians felt that the green line would influence decision-making. This was the only clinician to express enthusiasm for the AMS's capabilities to influence clinical decision-making:

U5: [. . .] that green line is really useful for me and I've not seen that before, maybe I should have but . . .

INT: No I think that's . . .

U5: That would be a real help to have in MDTs [multidisciplinary teams] and it's probably only been the last couple of years or even less than that we have had a pathology blood system that's shown us a graph of PSA and doubling time and that's made a massive difference in clinics to looking at things as sort of a visual aid, and this just takes it that step further, and I think in practice for me excluding the patient before they come in, that would be really helpful and in helping to explain things to some patients I think would also be useful, but it's biggest use would be for me as the clinician and for the MDT and follow up of patients would be great.

An examination of this clinician's response to the paper-based and AMS-based vignettes showed that the AMS did not drastically alter decision-making, but served a confirmatory purpose. The section of the interview that follows suggests that viewing the PSA values plotted as a graph ('black line') had the biggest impact on how he interpreted the vignette, whereas the green line provided further justification for initial reaction to the PSA trends:

INT (in relation to AMS presentation of Mr Evans' vignette): How is this different – having it in this format? So I think you mentioned you had a graph didn't you?

U5: Yeah, I would have the graph so I presumably would have the black line [showing Patient PSA values] there which actually. . . that in itself, looking at that black line now, compared to looking at a table – [. . .] would have worried me perhaps more than the figures that I looked at earlier. So I think the black line itself, which I have anyway, would have concerned me, but the additional information you have on there [showing predicted PSA for man with PCa] is really useful also because that's showing that he's going probably twice as fast as someone on active surveillance – so the additional information I've got there is ringing alarm bells to me so I think it would be useful.

All other clinicians who talked directly about the implications of the 'predicted PSA for a man with PCa' (green) line felt that this information alone was unlikely to change their clinical decision-making. These clinicians all emphasised that it is currently not possible to know the implications of following or digressing from the green line in terms of predicting long-term clinical outcomes for the patient, because long-term outcomes of patients on AS are not yet known. As a result, it was not possible to know the implications of following or deviating from specific PSA ranges. This was framed as the main limitation of the AMS:

*But what am I actually going to say to the patient?' Oh that means that you're typical of a man who's got cancer?' It's not telling me – this green line isn't of a cohort who go on to die of cancer, is it? Because you can say here, 'You're Mr Average prostate cancer active surveillance man.' But at the end of the day, we've still got to **tell** them, we've got to admit, really, that there's never been a good trial that shows that the outcomes of active surveillance, for the ones that progress, are as good as if they'd had treatment in the first place. Which is a kind of Achilles heel of the whole concept.*

U11

The clinician below was also aware of validation issues and current uncertainties surrounding PSA behaviour in PCa. Nonetheless, the green line was still branded as 'useful' in the sense that additional information was always welcomed, particularly in an uncertain field:

I was just thinking about your green line [showing predicted PSA for a man with PCa] when I was out. Because it's not validated it might just be encouraging us to do more biopsies that are not necessary. But I still think it's – it's useful. We need as many tools as we can get. We don't know whether any of them are useful at the moment, but hopefully in the fullness of time we will get that information.

U10

Relationship between lines: reassuring patients that they have 'low-risk' cancer

Clinicians felt that the AMS could be used to reassure men on AS by focusing on the relationship between lines. In particular, clinicians talked about demonstrating how close the 'predicted PSA for a man with PCa' (green) line was to the 'predicted PSA for a man without PCa' (blue) line, thereby demonstrating the relatively low-risk nature of a Gleason score of 6 cancer (see *Figure 9*). In other words, the AMS allowed for a visual 'normalisation' of a Gleason score of 6 disease:

And what that says more than anything else, that is saying that the insurance companies are right, Gleason 6 is not a proper prostate cancer, or rather, let me rephrase that, low-volume Gleason 6 isn't proper cancer, it is a condition that changes what the man in the street regards as cancer in a percentage of people. You know it's a pre-cancerous state or something or it's, you know it's not, cancer to the man on the street is something that will spread to the rest of his body and kill him. Low-volume low-rate prostate cancer rarely ever spreads and rarely ever kills anybody – and that backs it up.

U1

Four other clinicians gave similar views on how information about average PSA levels for men without PCa could be used to reinforce the idea that AS is the correct management route for men with low-risk PCa. It was noted that all clinicians who held this view were from one site, described by clinicians themselves as being particularly positive about AS:

Well I think what it confirms is actually that the men with prostate cancer – the green line [showing predicted PSA for a man with PCa] – is very close to the men without prostate cancer and therefore hence why they're the ones we probably don't really need to be worrying about too much as long as we just keep a close eye on them and maybe with the benefit of time we'll realise that we don't need to worry about them at all.

U4

Two of the five clinicians felt that the 'upper reference range for men without PCa' (orange) line was potentially reassuring for patients with low-risk cancer, simply because it could help to put their PSA into context. Clinicians expected patients to react positively to the idea that their values are still below a line associated with men without PCa:

In some ways that might be more reassuring for patients because you're saying actually you can have your PSA this high even if you don't have cancer, so and when patients only really think about PSA because they're not thinking about the biopsies and what we find when we examine them, they're worried about their blood tests, that orange line might actually be reassuring for patients to show them.

U4

Clinicians only interested in actual prostate-specific antigen line

Two clinicians could not see any benefits to having reference lines displaying average PSAs, explaining how the individual patient's data are all that one would be interested in analysing. This idea went hand in hand with the view that each patient's trends and rates of change, rather than absolute values, would be the underlying factor informing decision-making:

U2: Yeah, and again the (reference) lines here are helpful in terms of saying to the patient where he lies as sort of an average patient. But as I said to you the difficulty with these is that there is not a direct correlation always and these are individual patients.

INT: Yeah, of course.

U2: So it's helpful yes but you've got to qualify what you see on the screen with each patient individually.

One notable observation about clinicians' reactions to the AMS-based vignettes was the frequent tendency to focus on the hypothetical man's actual PSA values, without consideration to the reference lines. Clinicians sometimes needed prompting to focus on what value the reference lines added, rather than the intrinsic merits of having a graphical representation of PSA values generally:

INT: Would you find the cancer line helpful, do you think, so the men with prostate cancer and how their PSAs tend to change?

SPN1: Well I think a graph, plotted on a graph like that is better than looking at one on a table.

Importantly, most clinicians reported occasionally presenting PSA trends graphically in their standard practices. Consequently, although graphical presentation of PSA values was not standard practice for all clinicians all the time, it would not have been novel.

The active monitoring system as a whole: potential to influence clinicians' decision-making?

The AMS was generally thought to be unlikely to influence clinicians' practices or decision-making (with one single exception, presented above). Some clinicians were explicit about this but tended to qualify this view by talking about how the AMS could be helpful from a patient's point of view:

Probably and I think, my personal view is that it would be more useful for patients than it would be for me and I think it would help a discussion about what we would expect the changes in the PSA to be.

U3

Clinicians' main reasoning for expressing reluctance about the potential of the AMS to influence their decision-making related to their uncertainty surrounding the role of PSA measurements in predicting PCa behaviour. These contextual issues, relating to lack of evidence as to how reference ranges related to long-term clinical outcomes, were thus perceived to be the main limiting factors of the AMS:

We don't yet have good PSA data to show that it's a useful thing, so we're just making it up at the moment. And anything we can do to help improve that would be useful, which is why I tend to follow a rebiopsy rate anyway pretty much whatever the PSAs do.

U9

So, that might be helpful if you saw Mr Smith going along parallel with it, as a reassurance. Yeh, but, my own practice, mirroring that of the big experts, is to do a repeat biopsy occasionally, to avoid criticism as well as to help with evaluation.

U11

Clinicians' views on the AMS were gauged by asking whether or not they would want to incorporate it into their practice. Although responses were generally positive, the way in which this was conveyed suggested the benefits of doing so would be limited. One pattern to emerge was clinicians' suggestions that they had nothing to lose by having access to the reference ranges, thus framing this as an additional piece of information that could be of interest. There was a sense that having the AMS would 'not hurt', but its value had yet to be proven:

No I like this, this one, it's nice. I, I think . . . I don't think . . . if I had it, it'd be nice. But if I don't have it, it's still OK.

U7

Men's perspectives on data presented in the active monitoring system

Men's engagement with the active monitoring system

Four of the 20 men did not engage with the AMS at any level. These men tended to change the topic of discussion when asked for their thoughts on the AMS or produced closed responses. Further probing tended to result in comments about men being solely interested in what their clinician thought, and/or a disinterest in viewing statistics:

If you're shown the graph, and it, the doctor showed me this graph and said, 'This is you.' You're going to say, 'Should I be worried or have you got everything under control?'

P3

The above group of men formed a minority. Most men become absorbed into the AMS following explanation of what the lines represented. Five men found these data to be irrelevant to the way they thought about their condition and the care they received. These men understood what the AMS was showing but were solely interested in how their particular PSA values were behaving:

I mean effectively what this tries to bring is some sort of norm, isn't it. But for me it's still kind of at a personal level. I [. . .] It's almost like it's saying, ok, this is this, but the average Joe is this [. . .] I would probably base it on just on my own experiences of having the condition, I suppose, and what's happening with me. I think that would be more relevant than necessarily looking at what's happening on a graph to other people.

P15

The remaining 11 men saw some value in viewing their results in relation to the reference lines. Men valued the additional information the AMS provided and welcomed the possibility of viewing their own PSA results in context. For most men, this was purely an intellectual interest, but it still had potential to reassure.

Importantly, men found it difficult to be specific about how the AMS could be practically used but generally reported that they would not use the tool to make any decisions about their condition and its management. These views were congruent with their tendencies to defer decision-making to the clinician in their current AS follow-up (a key finding reported in *Chapter 6*). These views were more of a reflection of men's existing approaches to decision-making rather than a criticism of the AMS. There were, nonetheless, some elements of the AMS that attracted men's attention and promoted discussion. The data below are based largely on individual comments rather than on repeating themes.

Men's views on the upper reference ranges and predicted prostate-specific antigen for men without cancer (orange and blue lines, respectively)

A few men felt that the 'predicted PSA for a man without PCa' (blue) line had potential to be reassuring if a man's values ran close to this reference point:

I think, I think if I'd [...] ever been too concerned about it all, then I would've been far less concerned if I'd thought 'yeah well this isn't actually out of the ordinary.'

P10 (in relation to AMS presentation of Mr Smith's vignette where man's PSA values run just above the blue line)

The orange line was seen as potentially reassuring or alarming, depending on if figures fell below or above the line:

*Well I think it's very **interesting**, the 95th percentile line. Which shows, you know . . . that, in a way it's a reassurance, isn't it?*

P4 (in relation to AMS presentation of Mr Smith's vignette where man's PSA values fall below the orange line)

P4 [in relation to AMS presentation of 'Mr Jones' vignette where men's PSA values exceed the orange line]: Well, the fact that I was exceeding, you know, 95, you know – there's only, we're in this very tiny 5% aren't we?

INT: Yeah. I see. So the fact that it peaks above it?

*P4: So if I'm above **there**, oh heck, I really **am** in trouble.*

The fact that the orange line was visually the highest reference point may have played a part in shaping men's reactions:

*And (he's) nowhere near this **threatening** sort of red line [reference to the orange line] that, you know, you could get up to.*

P7

One man suggested that the mere spatial arrangement of lines would be sufficient to influence his interpretations of plotted PSA values. The fact that the orange line simply existed as the highest reference point ran the danger of making men complacent as long as they fell below this:

I'm not quite sure about this 95 line, whether that's helpful or not. [...] Because there's a danger that it distorts your thinking. Because you think 'Oh I'm a way off a risk, high risk level', when that might not be true. I don't know.

P14

One man considered the value of the orange line in relation to the blue line. The message that PSA values in men without PCa were so wide-ranging was deemed reassuring, as it further reinforced the man's pre-existing ideas about the limitations of PSA values when it came to monitoring cancer severity or progression. This man was keen to minimise the 'life intrusion' of routine PSA readings; this additional knowledge seemed to justify this stance:

Oo gosh. That's a really good one. I think – well it's made me think that actually made me even less anxious about PSA than I was. Which wasn't terribly . . . because it's more kind of random than I thought, in a way. Not random – that's the wrong word. More variable than I thought.

P13

Finally, similar to views expressed by clinicians, one man commented that the data on predicted PSA values for men without PCa (blue line) were irrelevant to men who were aware of their diagnosis:

Umm, I don't think that adds much [. . .]. Because, obviously, you would think that if you haven't got the cancer your PSA would be lower. Wouldn't it? So it shows it would be lower. So it's telling you what you know.

P12

Men's perspectives on the predicted prostate-specific antigen for a man with PCa (green line)

Similar to clinicians, the 'predicted PSA for a man with PCa' (green) line was the key reference point most men focused on, mainly because this was perceived to be the most relevant to them, given their condition. A recurring theme, discussed by seven men, was the reassurance associated with following the green line. This reassurance stemmed from the idea that their PSA values were not 'out of the ordinary' relative to other men in a similar position, thus giving them an opportunity to feel 'normal':

P7: You see I'd find that one [green line] the most reassuring of the three.

INT: What makes it reassuring on that one?

*P7: Because you're not far off the average. You know, you've **got** cancer. And you can't get away from that. Therefore you're looking at that green line. And you as an individual are pretty well normal – if there is such a thing in that circumstance.*

There was also a tendency for men to view the green line as an expected or anticipated trajectory, in a similar way to which some clinicians perceived this line. Following the green line was reassuring in that it gave a sense that a man's disease was behaving predictably:

Yes, well, I think seeing that I'd be less concerned. Because, you know, that's, we're on the expected trajectory.

P4

*Shows you the normal, standard, and you can see whether or not you are more or less following the same pattern. On the one you're showing, there, that chap is almost exactly following the standard anticipated, and he's still way below a level where it **can** be people without prostate cancer. I would say that would be reassuring, if that was myself.*

P6 (in relation to paper presentation of Mr Smith vignette)

One divergent case emerged in relation to men's thoughts on using the green line as a reference point. The individual below did not find the prospect of being a 'typical cancer patient' reassuring, possibly owing to a desire to distance himself from this label; he reported that he would need to fall below the predicted PSA for men with PCa to feel reassured:

P14: Because that's the line of the average cancer person, you think: 'That looks worrying that I'm on the line, you know,' [laughing] 'I'm a typical cancer sufferer.' So I would worry. Yes.

INT: OK. So following the green line you would associate with concern?

P14: Er yes. I mean that may be totally unscientific but . . .

Could the active monitoring system shape men's decision-making?

Only one of the 20 men interviewed felt that the AMS could influence his decision-making about his treatment. This man (P6) was unique in having taken control of decision-making throughout his AS experience, from the initial decision to opt for AS through to making the decision to refuse scheduled rebiopsies. This man was thus considered to be distinct relative to other interview participants.

The principal reason underlying most men's reactions to the AMS as a personal decision-making tool related to the attitudes uncovered throughout this chapter. Most men placed complete trust in clinicians to dictate their follow-up protocol, interpret their PSA findings and make decisions about when it was appropriate to consider radical treatment.

Men provided justification for why they felt that the AMS was of limited practical use to them, with responses falling under three categories: (1) difficulties with interpreting the AMS; (2) trust in clinicians' abilities to make the best possible decision for men; and (3) acknowledgement that decisions would not be made on the basis of PSA alone.

Men's difficulties interpreting the data presented in the active monitoring system

Three men were not confident in interpreting the information presented in the AMS:

I don't know, really. I think I'd, I'd need an expert interpretation as to exactly what it means. But, it is, you know, it's of interest, to see that, those lines.

P4

These men still engaged with the AMS, in the sense that they understood the general outline of what was presented. This discomfort in interpreting the information presented may have been related to a more general discomfort in reading graphs or statistics:

You're always . . . if you're shown the graph, and it, the doctor showed me this graph and said, this is you. You're going to say: 'Should I be worried or have you got everything under control?'

P3

Patients' overall trust in the clinician

Although some men tentatively interpreted the vignettes presented within the AMS, this was always expressed in terms of how they felt about the data (i.e. reassured or concerned). When probed to consider their next action, the response was, generally, to look towards the clinician for guidance. Thirteen participants explained that the decision to take action rested with their consultant:

No, I would still trust the people sat in front of . . . because I can't trust them then . . . you know, I'm not instantly a prostate cancer expert. So it'd be silly of me to start saying you got to do this and I demand that.

P7

Yeah, I think, I think it really would be done best with a surgeon. Or a, you know, specialist to discuss it. 'Cos I mean you could make any decision and you don't know whether it's the right one.

P16

Men perceived prostate-specific antigen as one of many indicators

In line with findings described earlier, five men expected treatment-related decisions to be based on a combination of factors, including DREs and – most importantly – biopsy results:

P9: You have to know the person. You have to know what the, the dialogue is with the consultant. And you know, because it's, they may well be, you know, healthy. They've had the biopsies, and the consultant's quite happy with it.

[Later] INT: If you were this gentleman and you were actually trying to decide what you want to do next, would these lines make a difference?

P9: Ah, no. Only if accompanied by a bad biopsy result or a consultant getting worried.

Men's views on potential uses for the active monitoring system

In general, men saw some value in the tool, even if this was simply to provide additional information. A number of men suggested more specific ways in which the tool could be usefully incorporated into a consultation, mainly for purposes of reassurance or explanation. Similar to clinicians' views presented earlier, one of these suggestions related to the early stages of joining an AS programme. In the example below, the man felt that there was value in using the AMS to explain that PSA rises in men with and without PCa:

*It might have been just useful just on the first consultation. Just to see that there are, you know, this does happen [. . .], it goes up with men in that sort of rate anyway, and what we might be looking for is inconsistency or a steady climb or whatever to worry about. But they kind of explain that, but without the graph. You know the graph just might help them to help **them** to show it a little bit easier.*

P9

Two men suggested that the AMS could be used as a visual aid to support the clinician's recommendation. In the example below, P12 emphasises the importance of the clinician retaining the decision-making power. He goes on to explain how the tool could justify the clinician's judgement:

P12: I know doctors will make mistakes – they are human beings. But I think I prefer a human being judgement [.] than a hard and fast rule that a doctor feels he has to apply because otherwise he's going to be sued sort of thing, you know? So I would say there is a risk in being too hard and fast with your decision tool. But it's, I mean, it is useful. I find as a man, the way I think, that's quite helpful. But I would not like it to be the last. . .

INT: Yes. I understand. Just to really pin down your views a little bit more, how do you think it could help you?

P12: I think it helps me to see the trend. This visible trend. It helps me to see where I am in relation to the average, and I think it would help me to come to terms with a decision. If, for example, the doctor said well, 'now, you see this is, you're obviously in a bad state.' Receiving that information might be easier when you can see that.

Clinicians' and men's views on presentation of data within the active monitoring system

Some clinicians expressed concern about the potential complexity of information presented in the AMS and the potential difficulties in trying to explain this within time-constrained consultations:

And you also have to be careful with patients because patients often don't understand the nuances of this, they don't understand an average is an average so you've got to be very careful how you explain these sorts of figures to patients, it can actually confuse them.

U2

In contrast, one clinician felt that the potential to add or remove reference lines would enable him to pitch the AMS at the right level, depending on his interpretation of patients' comprehension of graph/statistics:

To the clinicians, brilliant, so I think the ability to have, as you just did, add them and take them away, that would be a good thing. Because you might be just showing his line first and then showing the average man with cancer as an add on, a build on – that might be the way to do it.

U1

Men who provided comments on the potential benefits of the AMS offered little input into how the presentation could be improved. A few insights emerged from the researcher's observations of how men interacted with the AMS. For instance, two men found it difficult to see the green line, and one man viewed the orange line as red, a colour that can have 'danger-' or 'risk-'related connotations. If the AMS were to be developed in the future, these issues will need careful consideration.

Although eliciting views on presentation was generally challenging, a number of men were able to suggest additional information that they would have liked to have been displayed. These included CIs around the 'Predicted PSA for a man with PCa' (green) line, and the option to add in a line of best fit for patients' actual PSA values:

*P1: And if I, again, not very good on graphs, but if I were a statistician again or a graphologist [sic], whatever he is, I would want to take **mine** – that black line – entirely separately and I'd want to work out the average and reconfigure that, on that grid. **My** average. 'Cos this is my exact.*

INT: Oh I see, yes. Oh, OK.

P1: Now because there's a dip there, and these are two serious sort of dips. My mean line, when I work that out, may go somewhere like this [demonstrates].

INT: So you would want to do your – kind of – line of best fit?

P1: Yeah. I'd like to take my mean average [sic].

A handful of men expressed an interest in seeing reference lines that specifically indicated when they needed to take action:

*P6: I would have thought [. . .] rather than just the red line being sort of, y'know, saying that that's the highest of people without prostate cancer, maybe the red line could be, you know, a red line, if you **cross over** this red line, our experience would show that, you know, there is a [. . .] serious problem.*

*P12: Mm. There is another point, and that is, what would . . . without all this, at what the consultant says you'd better have an operation now. I wondered if you could put a line from **past** data, at the point when a decision was made. So I don't know how it would work exactly [. . .]. So basically he has the treatment when he's that much above deviation for that period. I don't know how you'd express that on the graph, but it. . . it [pause].*

INT: So almost, pooling together of the tipping points. . . .

P12: The tipping point. Exactly.

Conclusion

Clinicians and men found that the AMS provided useful additional information but questioned how much the tool would currently influence decision-making. Clinicians questioned the value and practicability of the AMS on the basis of its reliance on PSARRs, for which limited data were available regarding their relationship to long-term clinical outcomes, and the general uncertainties surrounding the role of PSA monitoring in AS protocols. Clinicians therefore did not believe that the system based on PSARRs would considerably change their current decision-making practices. Men's reliance on clinicians to direct their AS and future treatment trajectories resulted in little appreciation for how the system could be used as a personal decision-support tool, although this reflected men's general tendencies to defer treatment decision-making to the clinician.

Overall, clinicians and men valued the concept of having a set of reference ranges to support and direct AS decisions, in the sense that any additional information was helpful in this uncertain field. Although less of a dominant theme, both clinicians and men also identified the AMS as a useful tool for potentially providing reassurance through visual representation of where an individual's PSA value lies in relation to others with the disease and in relation to men without PCa. In this sense, the AMS was positively received as a tool that could be used to aid discussions and reassure, rather than influence, treatment decision-making.

Despite clinicians and men expressing doubts over whether or not the AMS could be practically used at present, the publication of long-term patient outcomes and how these relate to PSA behaviour may change this. Although not dominant themes, a number of useful suggestions for improvement to the AMS emerged from patients' and clinicians' accounts. These included presenting a line to denote when the individual man should consider alternatives to AS, and the addition of the option to add a line of 'best fit' through a man's individual PSA values. Including lines that prompt when men/clinicians should consider alternatives to AS might be made possible by publication of outcome data from current trials (see *Chapter 8*). Options for more pragmatic changes include further exploration of which colours may be best used in the AMS. This is based on the researcher's observation that some men had difficulties distinguishing some of the reference lines.

Chapter 8 Randomised controlled trials of active monitoring strategies

Aim and background

This component of the project aimed to design a RCT to evaluate the effectiveness of the most acceptable 'AM' instrument compared with existing AM methods.

There have been few RCTs comparing conservative management with radical treatment for clinically localised PCa. The PIVOT recently found no difference between observation (i.e. WW) and radical prostatectomy for either all-cause or PCa mortality after at least 10 years of follow-up,¹¹ although the cohort was relatively elderly (mean age of 67 years). SPCG4¹² found that radical prostatectomy reduced PCSM compared with observation (i.e. WW) among men who had been diagnosed clinically as opposed to screen-detected in PIVOT. The ProtecT RCT is the first to compare AM with radical treatment.^{14,15} If results of this trial (to be published in 2016) find evidence that AM is an effective option for men with clinically localised PCa, then a RCT comparing the effectiveness of the AMS against other monitoring or AS strategies, such as the use of PSA kinetic measures, would be of considerable importance. Currently, however, without the ProtecT trial results, and with the relative immaturity of the existing cohorts of men on AM/AS protocols resulting in few events in terms of metastases or death (as shown in *Chapter 2*), the evidence as to the optimal protocol for non-radical treatment is not known. Here, therefore, we give an overview of some of the issues surrounding potential trial design of AM/AS strategies, including issues identified by this study, and outline several potential trial designs which could then be tailored for the definitive design, depending on results from the ProtecT trial in 2016.

Difficulties related to new randomised controlled trials of active monitoring strategies

Recruitment and design

The ProtecT trial successfully gained consent from over 1600 men to be randomised to AM, radical prostatectomy or radiotherapy. However, as the failure of the START (Phase III Study of Active Surveillance Therapy against Radical Treatment in Patients Diagnosed with Favourable-Risk Prostate Cancer) trial¹⁰⁸ demonstrates, recruiting enough men who are willing to be randomised to such different approaches remains an extremely difficult task. Lessons learned from the ProtecT trial would need to be implemented.¹⁰⁹ Recruitment of men to a trial of AS/AM would need to take place following PSA testing in a healthy population, or following ad hoc testing in clinical practice. A key issue in recruiting to a PSA-tested population is the huge numbers required – over 100,000 tested in the ProtecT trial, for example.^{14,15} The alternative is to recruit from the pool of all new cases of clinically localised PCa who according to the latest NICE guidance,⁶⁸ should be offered this option if they have low-risk disease or higher-risk disease but are unwilling to have radical treatment. Men would need to accept an initial conservative approach to their newly diagnosed cancer and to give consent to be randomised to one of several AM/AS strategies.

A decision about whether or not the study would involve randomisation at the cluster (clinic) or individual level would need to be made based on the AM/AS strategies under investigation. For example, a trial comparing monitoring with the AMS and an alternative monitoring strategy (e.g. current practice) would need to be cluster randomised, because there would likely be contamination between interventions if members of the same clinic were randomised to different monitoring strategies. Blinding of clinicians and patients would not be possible in a trial of different monitoring strategies, given the obvious distinctions between potential interventions, but blinding of outcome assessment would be considered carefully.

Pre-trial and ongoing training of doctors or nurses responsible for AM/AS check-ups would be vital in any such trial. The interpretation of PSADT, PSA_v and the PSARR graphs would need to be clear and consistent.

Eligibility

An issue for any trial will be to decide on the eligibility criteria for enrolment. The eligibility criteria for the ProtecT trial mean that it will be difficult to reach conclusions in the higher-risk groups (i.e. men with Gleason scores of 3 + 4 or 4 + 3). However, as worldwide AS studies accrue men and time in follow-up, the safety of allowing such men to enrol in AM/AS may be established.

Several studies have published eligibility criteria along with validation, the most popular of these being the D'Amico criteria,^{110,111} where low risk is: PSA < 10 ng/ml, Gleason score at most 3 + 3 and stage up to T2a; intermediate risk is PSA between 10 ng/ml and 20 ng/ml, Gleason score up to 4 + 3 and stage T2b; high risk is PSA > 20 ng/ml, Gleason score of at least 4 + 4 and stage T2c or T3a. Recently, there has been a proposal to use a single early PSA test to stratify men as low, intermediate or high risk.¹¹² If the ProtecT trial demonstrates that AM is an effective option for low-risk cancer, then a trial might be needed to establish whether or not it is also effective for those with intermediate-risk localised cancer.

Recent developments in MRI¹¹³ and potential genetic variants associated with lethal PCa¹¹⁴ may lead to better stratification of those men with PCa who will not experience progression to life-threatening disease. It might be that, in future, more precise stratification might be employed to determine eligibility to enter a trial comparing different management strategies.

Outcomes

The results of this project, including the systematic review, suggest that PCSM is rare in the short term for patients on AM/AS. However, PCSM or metastases are the most important primary outcomes by which to compare any monitoring strategies and should form the primary clinical outcome for future trials. Either a large sample size or long-term follow-up (or both) would be required to produce enough events to allow for any risk-group comparisons. However, if the ProtecT trial demonstrates that AM is effective, then other primary outcomes could be considered when comparing AM/AS strategies, such as the number of men remaining on AM/AS over time or to a specific age, adherence to monitoring protocols, costs and ease of implementation in the NHS, patient satisfaction and quality of life. The chosen primary outcome would have a crucial impact on the design issues of sample size and length of study. Repeated biopsy results are another potential clinical outcome that could be considered, but the sampling procedures (e.g. number of cores used) introduce a large amount of measurement error with either under- or oversampling of the prostate. For example, in the PRIAS [(Prostate Cancer Research International: Active Surveillance) see www.prias-project.org] study 2494 men were diagnosed with PCa and followed AS; of the 1858 repeat biopsies taken in the study, 687 found no cancer.⁸⁹

Potential trial designs

Active monitoring system

A trial comparing the use of the AMS with use of current PSA kinetic thresholds and methods might be of interest to test the acceptability of the AMS, but the issues around what form of AS or AM would be included remain. Our qualitative research has revealed that a minority of patients and clinicians express concerns about the complexity of using a system or graph for PSA monitoring, although most could see some potential for reassurance through using the AMS. However, consistent interpretation of the PSARRs presents a potential problem for nurses, doctors and patients. A short training session for nurses and/or doctors would be needed before commencing any trial to ensure accurate delivery of the system and to minimise interobserver bias in interpretation and informing men. It would be helpful to have objective criteria that are strictly adhered to, such that once a single PSA value falls above the 95% reference range, the nurse/doctor recommends re-evaluation of the cancer, although whether or not such clear criteria will ever be discovered remains uncertain. Furthermore, in the interest of time, nurses/doctors would need

adequate training to explain clearly and concisely what the graph is telling the patient, such that consultation time is not wasted. Interviews with patients have made clear that these graphs should be shown to the men in a clinical setting and not sent out to men to interpret without guidance.

Qualitative research has suggested that PSADT and PSA_v are sometimes calculated crudely and in a variety of ways in a clinical setting, leading to variable results. There are also many different methods of calculation, as have been discussed previously^{29,31} and, thus, a computer is needed at every visit to compute PSA kinetics consistently. It would be necessary to identify the optimum method and to ensure that it is applied consistently by suitably training those doctors/nurses in this arm of the trial.

This trial would require a cluster-randomised design, with clinics or GP surgeries as the unit of randomisation. The primary outcome could be the number of men remaining on monitoring at a median follow-up of, for example, 5 years, with a potential further follow-up depending on the rate of management change in the trial. Adherence to the monitoring system, quality of life and satisfaction with monitoring could be secondary outcomes. Reasons for leaving AM would be recorded but not limited to a single reason, given that several distinct factors could lead to a decision to seek radical intervention.

Active monitoring compared with active surveillance

Although ongoing studies with AM and AS may produce useful comparisons, they will not be able to establish as strong a conclusion as a RCT comparing scheduled repeat biopsy with using biopsy only when radical treatment is being considered. A trial comparing AM with AS (i.e. men randomised to have a rebiopsy only when triggered or at scheduled intervals) would determine whether or not this extra procedure needs to be regularly performed to increase the effectiveness of monitoring. *Chapter 2* found several frequencies of rebiopsy even among the 16 AS studies, ranging from annual biopsies to biopsy at 1, 4 or 7 years.

The patients would need to be selected from participating clinics and randomised to AM or AS. A cluster randomised design may be necessary in this trial to avoid anxiety due to perceived reduced testing in the AM arm and possible anxiety due to overtesting in the AS arm, as well as contamination with men in the other arm of the trial.

The primary outcome could be the number of men remaining on AM or AS at a median follow-up of, for example, 5 years, with potential extension depending on the rate of management change. Secondary outcomes would be adherence to AM or AS, patient satisfaction, number of complications from rebiopsy, quality-of-life score and reasons for seeking radical intervention.

Active monitoring for higher-risk or locally advanced prostate cancer

If the results of the ProtecT trial show that AM is an effective option for men with clinically localised T1–T2 PCa, then a future trial could investigate whether or not AM/AS is effective in men with higher-risk or more advanced T3a disease. This option for a RCT would be difficult to launch currently because of concerns about the levels of progression likely to be experienced, as well as the NICE guidance advising men in these groups to have radical treatment.⁶⁸

Nurse-led active monitoring

Qualitative research within the ProtecT trial has shown that both doctors and patients find nurse-led monitoring acceptable. As such, a potential trial that could be undertaken now could compare nurse-led clinics with usual care involving consultant urologists. Some patients in this study have indicated that appointments with urologists were often so short that they wondered if there was a need for them to be involved, and travel and parking considerations were also raised. If this trial were conducted, men randomised to monitoring by nurses would need the support of consultants either at fixed time periods (perhaps annually) or when any change in disease status or anxiety required a clinical review. A nurse-led AM/AS programme, such as the one employed in the ProtecT trial, could lead to NHS savings from freeing up consultant time. A full economic evaluation would need to be undertaken, and cost could be a key

outcome. The primary outcome in such a trial could include satisfaction with and adherence to monitoring, with quality-of-life score and number remaining on monitoring as secondary outcomes. The randomisation could be performed at the individual level, with men visiting either nurses or consultants at that clinic, or it may need to be a cluster design.

Summary

The precise design of trials of AM/AS strategies needs to wait for the results of the ProtecT trial in 2016 to provide information about the patient risk group and the effectiveness over the long term for patients with particular clinical and demographic characteristics. We have suggested four potential trial designs, which would add considerably to the evidence required to place AM at the forefront of care for men with recently diagnosed clinically localised PCa – including one that could be launched soon. Once data from the ProtecT trial become available, a comparison of PSARRs, PSA kinetic measures and the PSA protocol used in ProtecT could determine whether or not PSARRs lead to a more effective protocol for AM/AS in the UK population. Preliminary results from this project have shown partial evidence that PSARRs are an improvement on PSA kinetic measures for monitoring PSA over the course of AM/AS. However, all trial designs are limited by the lack of robust trial evidence about the effectiveness of AM/AS as a strategy, necessitating the wait for the ProtecT trial results.

Chapter 9 Conclusions and recommendations

Summary of findings

The systematic review (see *Chapter 2*) showed that there is little evidence of any consensus on many aspects of the approach to AM or AS, such as inclusion criteria, methods of monitoring and triggers for change of management. However, our meta-analysis (see *Chapter 2*) suggests that only 84 changes of management occur per 1000 person-years (95% CI 61 to 106 changes of management per 1000 person-years). We found that this rate was associated with Gleason score and whether or not the study used repeat biopsy testing in its monitoring schedule. In *Chapter 3*, a model for the PSA monitoring system was developed using data from the ProtecT study AM cohort. The evidence from *Chapter 4* suggests that this model captures the trend in age-related PSA change in modern AM/AS cohorts in both the UK and the USA. In *Chapter 5* there was some evidence to suggest that PSADT and PSARRs could help discriminate men who will die from PCa from those who will not. However, there were mixed results when comparing the markers, such that no strong recommendation for any can be made here. The validation of the AMS to alert men to pathological progression has been hampered by low numbers of clinical events in the immature cohorts available. The ProtecT trial will provide evidence about the effectiveness of AM for men with clinically localised PCa in 2016. Once these results are available, the effectiveness and cost-effectiveness of the AMS can be evaluated in a RCT, perhaps in conjunction with nurse-led clinics.

The qualitative work undertaken (see *Chapters 6 and 7*) explored the acceptability of the PSARRs developed in previous chapters in the context of current clinical practice. PSARRs were incorporated into a computerised system (the 'AMS'), and shown to clinicians and men involved in providing and receiving AS in NHS hospitals in England. Given that the systematic review (see *Chapter 2*) established that there was no consensus on AS processes, *Chapter 6* focused on mapping out current AS provision in clinical practice in these hospitals and capturing the perspectives of clinicians and patients in relation to this uncertain field. This work provided an important foundation for understanding how the AMS might be integrated into clinical care in the future. *Chapter 7* focused on clinicians' and patients' views about the AMS, with a view to understanding whether or not the system had potential to contribute to management and/or influence decision-making (and if so, how).

Chapter 6 considered men's understanding of and views surrounding AM/AS protocols, including thresholds for moving off AM/AS. These topics had not yet been investigated in depth, although men's general views on AS as a treatment option had been explored in a handful of studies outside the UK (see *Chapter 1, Qualitative research in active monitoring/active surveillance*). Investigating patients' and clinicians' perspectives of current AS provision in UK practice was an important foundation for understanding how the AMS may be integrated into practice. *Chapter 6* presented the various ways in which uncertainty in AS manifested at the level of clinical practice, and the ways in which clinicians and patients responded to these uncertainties. The uncertainties reflected those in the systematic review and surrounded clinicians' decisions regarding which patients should be eligible for AS, how AS follow-up should be undertaken and the triggers that should prompt consideration of radical treatment. Uncertainty also existed in relation to measurement methods used to characterise PCa, as clinicians pointed out limitations with the clinical parameters they currently used. In response, much of AS decision-making appeared to be a process of considering multiple clinical factors, without clear thresholds. The inherent uncertainties in the field were reflected in our finding that clinicians' reported practices varied within and across sites. The existence of protocols at some sites appeared to be associated with greater consistency in reported practices, although clinicians were still explicit about the need to deviate from protocols depending on particular circumstances.

Overall, clinicians were mindful of decisions that they were sometimes required to make without a solid evidence base. Patients, however, appeared content to trust their clinician as the principal driver in decision-making within AS. There was a tendency for patients to look to their urologists to interpret their clinical progress on AS and to indicate when it might be more appropriate to consider radical treatment. Patients tended not to be aware of all the uncertainties that clinicians discussed in the interviews, but were mindful of the limitations of PSA testing as a prognostic tool. This awareness by men contributed to their understanding and acceptance of multiple follow-up tests – including rebiopsy – in AS programmes, although some patients used repeatedly steady PSA values as justification to refuse rebiopsy. Some men also questioned the necessity of having to visit the hospital to see a consultant when PSA levels remained stable. This suggests that individual patients weighed up the personal costs and benefits of accepting different investigational procedures.

Chapter 7 showed that men and clinicians responded positively to the availability of additional information provided by the AMS and its potential to provide reassurance, but all expressed doubts over whether or not the tool currently had the potential to influence current decision-making. The findings from *Chapter 6* help to elucidate the reasons underlying this; patients tended to hand over decision-making power to clinicians, and clinicians were reluctant to base clinical decisions on a tool that considered PSA values alone. This related to wider uncertainties surrounding the role of PSA in AS and gaps in the evidence pertaining to how PSA values translated into clinical outcomes. Findings from *Chapter 6* established how decision-making involved weighing multiple sources of information and reaching a decision acceptable to both parties.

Despite beliefs that the AMS as a single measure was unlikely to influence decision-making, a subset of patients and clinicians recognised how the tool might be incorporated into practice as a visual aid to provide reassurance by placing PSA values into context (i.e. in relation to men without PCa, or similar men in their position). However, this scope for reassurance would clearly depend on how the individual patient's PSA levels were behaving in relation to the reference lines that could be shown, as well as on evidence about the long-term effectiveness of AM/AS strategies.

Discussion

The number of AS studies is increasing worldwide, and, with a continued lack of consensus about key principles, it is important to investigate variation between studies. There is accumulating evidence that AM/AS can be a safe option for men, including the small number of cases of metastases and PCa death in contemporary cohorts (see *Chapter 2*) and the evidence from the US PIVOT trial of the lack of benefit of surgery compared with WW.¹¹ However, the AM/AS cohorts are still immature, with short follow-up periods and relatively few events. When they do mature, the highly selective patient groups, diversity of eligibility and monitoring strategies, and wide variations in triggers for management change mean that analysis of these cohorts will be prone to bias, and it will be challenging to compare strategies and form effective policies. The National Institutes of Health (NIH)-funded PROMISS (PROstate Modelling to Identify Surveillance Strategies) study, combining several cohorts and modelling, will be important in interpreting these results. The dilemmas inherent in PCa screening are also present in the evaluation of AM/AS, particularly in determining strategies of monitoring that will enable the majority of men with low-risk disease to avoid radical treatment but that will identify those in whom such radical treatment would be beneficial. The ProtecT trial, comparing an AM programme without scheduled rebiopsy with radical prostatectomy and radical conformal radiotherapy will report in 2016.^{14,15} Other trials, particularly comparing different AS/AM protocols, are needed to provide unbiased comparative evidence about the most effective and acceptable strategies for men and clinicians. The NICE guidelines for AS are based on the ongoing RMH study⁸⁵ which has been analysed in this report. Although it is commendable to attempt to generalise an approach to AS, there exists no evidence on the 'correct' or consensus approach to AM/AS. As demonstrated in *Chapter 2*, we are only at the stage at which we can compare different AM/AS study designs and identify aspects of the AM/AS protocol that impact upon some outcomes. This area of research is still in its infancy.

Studies of men with low-risk PCa have indicated that AS is safe (i.e. there is a low rate of death from PCa for men on AS).⁸² However, there is no consensus on a detailed definition of 'low risk', and many cohorts appear to be limiting AS to those in an 'ultra-low-risk' group. Until randomised trials comparing different treatments publish, evidence is limited to institutional comparisons. Even when these trials report their results, the evidence for which subgroups should be offered which treatments will still be limited. Thus, future studies need to state definitions of low-risk groups clearly, and comparisons between studies of different designs may shed light on this question about who is most suited to AM or AS.

There are limitations in the statistical analyses carried out in a number of studies found in *Chapter 2*. The use of simple comparative techniques such as *t*-tests and Mann–Whitney *U*-tests to find differences does not allow for potential confounding. Extended Cox models that allow for time varying covariates such as PSA were not used by any analysis, thus wasting information provided by serial PSA results. The estimation of the effect of Gleason score or PSA, for instance, on cancer progression is the main question of interest in many analyses. However, because progression is often defined by thresholds of Gleason score or PSA, it is difficult to untangle and quantify whether or not these effects are 'true' progression. A gold-standard variable for proven disease progression is needed but, unfortunately, is not currently available. With longer follow-up, stronger outcomes of metastases or PCSM will be more frequent and, thus, more reliable.

One of the most pressing issues in AS is the reliability of PSA as a predictor of disease progression.¹¹⁵ PSA has high day-to-day biological variation,¹¹⁶ caused by many factors, such that a single high measure may not be necessarily indicative of a worsening condition. Depending on the method of calculation, PSADT or PSA_v could be sensitive to such fluctuation, thus suggesting 'progression' when it may not be warranted. Some studies require a confirmatory PSA test when such a result occurs. Few alternatives to PSA are available. A subanalysis of the JH cohort found that proPSA (a form of PSA which is increased in the blood of a man with PCa) was associated with time to an unfavourable biopsy result for men on AS.¹¹⁷ Recently, the use of some genetic biomarkers for PCa detection,¹¹⁸ distinguishing lethal PCa¹¹⁴ and predicting post-radical prostatectomy recurrence¹¹⁹ have been proposed, but definitive results have yet to be found for use in AS. MRI and magnetic spectroscopic imaging (MRSI) are being used to follow up men in some AS cohorts.¹²⁰ A recent systematic review of novel tools for AM/AS has recommended that prospective studies should include and evaluate these new markers, such as imaging.¹²¹

Any instrument aiming to predict outcome for men with low-risk PCa will need to be calibrated to the population and specific clinical need. For example, in a setting in which only ultra-low-risk men were offered AM, clinicians may decide to opt for a higher reference range (e.g. 99%) in order to raise the specificity and avoid unnecessary treatment. However, were a new radical treatment for low-risk PCa to be developed, which had fewer side-effects than either surgery or radiotherapy, clinicians and patients may prefer to lower the reference range, ensuring that more men would be treated unnecessarily but balancing this against the decreased risk of missing a progressing cancer. Similarly, the instrument would need recalibrating between populations where screening was common, uncommon and rare.

As already discussed, the quantitative findings were severely hampered by a lack of clinical outcomes or events. Furthermore, the SPCG4 cohort, which contained data on onset of metastases and PCSM, was a WW arm of a larger RCT and recruited in the pre-PSA era. Given the intent to use the AMS prospectively, it would be ideal to validate it using an ongoing AS cohort. We have attempted to resolve this issue by comparing the AMS with PSA kinetic measures only in those SPCG4 men with a diagnostic PSA < 20 ng/ml and Gleason score of 6 or less. Another limitation is a lack of clinical covariates, such as number of biopsy cores positive for cancer or tumour volume, which may have allowed a more comprehensive AMS to be developed. However, as shown by the systematic review, not all studies collected these regularly, and a simpler set of covariates (e.g. age, PSA, Gleason score and T-stage) was used to ensure comparability across studies.

A strength of the quantitative part of this study is that large, independent cohorts from different countries and years were available for model development and validation. The model was developed using data from over 7000 PSA tests (see *Chapter 3*), and we have attempted to predict over 15,000 PSA test results (see *Chapter 4*). Our data were from both US and UK populations and traversed two eras of PCa detection: the clinically presenting men from the early 1990s and the PSA-detected men of the 2000s. We have used a comprehensive modelling strategy and validated our models both in terms of prediction of PSA and in terms of predicting clinical outcomes.

The qualitative findings were limited through reliance on participants' reports of practices and recollections of events rather than observations of actual interactions. As such, data may have been subjected to respondent biases and tendencies to offer socially desirable responses. This is particularly relevant to clinicians, who may have felt their practices were being subjected to scrutiny. Men's accounts were also at risk of similar biases, including issues of recollection bias. This was especially pertinent to patient interviews, as data often constituted narratives of what had occurred in previous discussions with clinicians. These data will also have been affected by patients' interpretations, which may differ to those of clinicians and/or researchers undertaking observations of actual interactions.

The qualitative study samples were limited on a number of counts. Only sites in the south of England were sampled, due to pragmatic and logistical reasons. The sample of men interviewed was limited by sampling low-risk patients. Men at the periphery of AS eligibility criteria may have expressed different ideas in terms of reassurance and involvement in clinical decision-making. The sample of men interviewed was selected by clinicians. As such, the sample may have reflected patients that clinicians deemed to be appropriate for purposes of a research study, whether this was conscious or not.

Although the sample size was small, we are satisfied that data saturation was achieved with respect to evaluating the AMS tool. However, some of the ideas inductively derived in *Chapter 6* pertained to broader topics in relation to conducting and receiving AS, many of which could have benefited from further investigation with continued sampling. Therefore, sample numbers were, in part, influenced by practical considerations (e.g. study timelines), although we are confident that this did not hamper the primary objectives of the qualitative study.

A strength of the qualitative study was that it was able to provide a novel insight into patients' and clinicians' perspectives on AS in practice. Although the primary aim was to evaluate the AMS, the qualitative study was strengthened by incorporating the broader objective of investigating views and experiences surrounding current AS practices. This allowed for multiple themes relating to uncertainty to emerge inductively from data. These ideas have opened up new avenues of enquiry for future research, as reported below.

Research recommendations

Modelling the relationship between prostate-specific antigen and prostate cancer outcomes

We recommend that studies in this area share data in order to provide the best evidence for future management of men with clinically localised PCa. The PSA era began in the early 1990s, and the earliest ongoing study of AS began in 1995. In our systematic review, just eight of 7111 men followed for roughly 4 years had died from PCa. It will take some time before the worldwide cohorts of men on AM/AS mature, and, even then, the number of deaths from PCa may be small because men in these cohorts die from other causes. Thus, the best examination of the effect of PSA change on PCSM or metastases will be through combining data from various research cohorts, using meta-analysis and advanced modelling techniques, including, for example, the US NIH-funded PROMISS study now under way.

Reporting in active monitoring/active surveillance studies

The systematic review in *Chapter 2* found little consensus on study design between 22 cohorts of men followed by AM/AS. Furthermore, there was little consensus on what to report, despite the similar aims of these studies. For instance, although the reasons men changed AM/AS to radical treatment were not provided by each study, this information is likely to be crucial in future comparisons (through meta-analysis) between studies. Given the low short-term PCSM rate found in the systematic review, now would be the ideal time for ongoing studies to agree on a core outcomes set (as in clinical trials), whereby each study records instances of metastases and PCSM, for instance. As discussed in *Chapter 2*, it is likely that future comparisons to decide on the optimal strategy for AM/AS will be carried out through meta-analysis. Thus, it is crucial that a wide range of studies is undertaken and that they report design, triggers and results in as similar a way as possible. Future studies should also collect data on those who choose to leave AM/AS and the reasons for this decision, and should continue to follow these patients. Data on quality of life will be important for future AM/AS studies, along with an increasing focus on imaging, biomarkers and genetic information.

Changing clinical environment

The developments in functional magnetic resonance imaging (fMRI) and other imaging technologies may change the diagnosis and risk-stratification of men with PCa. These developments, and others in relation to clearer understandings of genetic risk, will need to be pursued in future research. We would recommend that cohorts of men with PCa build up a biorepository allowing retrospective measurement and evaluation of any new biomarkers. This will increase power to detect if new biomarkers are useful for monitoring men with clinically localised disease.

Future qualitative research in active monitoring/active surveillance

In terms of qualitative research, future work may consider supportive interventions to help clinicians and men cope with the various uncertainties in AS. Clinical centres may want to consider the use of formal protocols, with the opportunity to reduce practice variation and increase clinician confidence and patient reassurance. These ideas were not pursued in the current project, but our observations of subtle differences between sites with and without protocols suggest that this may be an important avenue to consider in future work.

The research in *Chapter 6* suggested that there may be greater emphasis on patient choice and preference where clinical uncertainty is heightened, although this was generated as a hypothesis rather than a finding. Future work may investigate these notions more fully and could consider the use of observational techniques to capture clinician–patient interactions, particularly at key decision-making points in patient pathways. We uncovered evidence to suggest that there were differences in how central the patient's role in AS decision-making may be, although these ideas will require more thorough investigation. Future research could explore the impact of centre protocols on levels of patient involvement and how optimum patient involvement can be achieved for individual patients regardless of centre protocols.

Qualitative interviews with men on AS suggested that the value of outpatient clinical appointments would benefit from further investigation. In particular, future work may consider the content and duration of appointments. Although only raised by a small number of individuals, there was a suggestion in this study that routine face-to-face appointments with urologists may not be deemed essential by men and their partners, although this appeared to depend on the content and duration of appointments. Where outpatient appointments are unlikely to call for clinical examination or discussion of change in management, there may be options to promote non-face-to-face forms of communication [i.e. telephone, Skype™ (Microsoft Corporation, Redmond, WA, USA)], or appointments led by specialist nurses. A further option may be to conduct outpatient appointments in a targeted, rather than routine, manner. These different options could have considerable implications for NHS resources and patients alike, especially given the increasing prevalence of men on AS.

Finally, future qualitative work could reconsider the application of the AMS once more is known about the effectiveness of PSA monitoring and the clinical implications of PSA values/behaviour. These gaps in evidence were the primary reason for clinicians' reluctance to view the AMS as being capable of influencing clinical decision-making at present. Ultimately, the development of AS and AM programmes, and systems such as the AMS, await the publication of ongoing research including the ProtecT trial.

Acknowledgements

We would like to thank:

- all men, clinicians and nurses who gave up their time to be interviewed as part of this project
- Mr Simon Brewster, Mr Simon Evans, Professor David Gillatt and Mr Mark Speakman, all of whose contributions made the interview substudy possible
- Mr Michael Davis for preparing a data extract from the ProtecT study
- Professor H Ballentine Carter for contributing data from the JH AS cohort
- Professor H Ballentine Carter, Dr Luigi Ferrucci and Dr Jeff Metter for contributing data from the BLSA
- Dr Chris Parker for contributing data from the RMH AS cohort
- Dr Anna Bill-Axelson for contributing data from the SPCG4 WW cohort
- Professor Ruud Bosch for contributing results from the Krimpen study
- Professor Jeremy Taylor and Professor Ruth Etzioni for ongoing discussions of statistical modelling.

Contributions of authors

Andrew J Simpkin drafted *Chapters 1–5, 7 and 8*, along with the quantitative portions of the abstract, plain English summary, scientific summary and the appendices. He carried out all quantitative analyses, collated comments and redrafted the report.

Leila Rooshenas conducted the qualitative interviews and analyses and drafted *Chapters 6 and 7*, along with qualitative portions of the introduction, abstract, plain English summary, scientific summary, conclusions and appendices.

Julia Wade contributed to the study design, contributed to and supervised the qualitative data collection and analyses, and contributed to critical revision of manuscript drafts.

Jenny L Donovan contributed to the overall study design, data collection (for the ProtecT trial), designed the qualitative study, and contributed to the data interpretation and critical revision of manuscript drafts.

J Athene Lane, Richard M Martin, Freddie C Hamdy, David E Neal contributed to the study design, data collection (for the ProtecT trial), data interpretation and critical revision of manuscript drafts.

Peter C Albertsen contributed to the study design, data collection (for UCHC) and critical revision of manuscript drafts.

Lars Holmberg contributed to the study design, data collection (for SPCG4), data interpretation and critical revision of manuscript drafts.

Kate Tilling led the project, and, with **Andrew J Simpkin** and **Leila Rooshenas**, had full access to the data used, and takes full responsibility for the integrity of the results and accuracy of the summaries presented here. All authors critically reviewed the manuscript.

Kate Tilling and **Chris Metcalfe** contributed to the study design, supervised the quantitative data analyses, contributed to the quantitative data interpretation and critical revision of manuscript drafts.

Publications

Simpkin AJ, Metcalfe C, Martin RM, Lane JA, Donovan JL, Hamdy FC, *et al.* Longitudinal prostate-specific antigen reference ranges: choosing the underlying model of age-related changes [published online ahead of print October 9 2013]. *Stat Methods Med Res* 2013.

Simpkin AJ, Tilling K, Martin RM, Lane JA, Hamdy FC, Holmberg L, *et al.* Systematic review and meta-analysis of factors determining change to radical treatment in active surveillance for localized prostate cancer. *Eur Urol* 2015;**67**:1006–8.

References

1. Cancer Research UK. *Cancer Stats*. 2012. URL: <http://info.cancerresearchuk.org/cancerstats/> (accessed 16 February 2012).
2. Sakr W, Haas G, Cassin B, Pontes J, Crissman J. The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. *J Urol* 1993;**150**:379.
3. Howlander N, Noone A, Krapcho M, Neyman N, Aminou R, Waldron W, *et al*. *SEER Cancer Statistics Review, 1975–2008*. Bethesda, MD: National Cancer Institute; 2011.
4. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;**62**:10–29. <http://dx.doi.org/10.3322/caac.20138>
5. 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. *N Engl J Med* 1991;**324**:1156–61. <http://dx.doi.org/10.1056/NEJM199104253241702>
6. Hoffman RM. Screening for prostate cancer. *N Engl J Med* 2011;**365**:2013–19. <http://dx.doi.org/10.1056/NEJMcp1103642>
7. Catalona WJ, Ramos CG, Carvalhal GF, Yan Y. Lowering PSA cutoffs to enhance detection of curable prostate cancer. *Urology* 2000;**55**:791–5. [http://dx.doi.org/10.1016/S0090-4295\(99\)00602-0](http://dx.doi.org/10.1016/S0090-4295(99)00602-0)
8. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Ciatto S, Nelen V, *et al*. Screening and prostate-cancer mortality in a randomised European study. *N Engl J Med* 2009;**360**:1320–8. <http://dx.doi.org/10.1056/NEJMoa0810084>
9. Frankel S, Smith GD, Donovan J, Neal D. Screening for prostate cancer. *Lancet* 2003;**361**:1122–8. [http://dx.doi.org/10.1016/S0140-6736\(03\)12890-5](http://dx.doi.org/10.1016/S0140-6736(03)12890-5)
10. Dahabreh IJ, Chung M, Balk EM, Winifred WY, Mathew P, Lau J, *et al*. Active surveillance in men with localized prostate cancer: a systematic review. *Ann Intern Med* 2012;**156**:582–90. <http://dx.doi.org/10.7326/0003-4819-156-8-201204170-00009>
11. Wilt TJ, Brawer MK, Jones KM, Barry MJ, Aronson WJ, Fox S, *et al*. Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med* 2012;**367**:203–13. <http://dx.doi.org/10.1056/NEJMoa1113162>
12. Bill-Axelson A, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C, *et al*. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 2011;**364**:1708–17. <http://dx.doi.org/10.1056/NEJMoa1011967>
13. Holmberg L, Bill-Axelson A, Helgesen F, Salo JO, Folmerz P, Häggman M, *et al*. A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. *N Engl J Med* 2002;**347**:781–9. <http://dx.doi.org/10.1056/NEJMoa012794>
14. Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, *et al*. Prostate Testing for Cancer and Treatment (ProtecT) feasibility study. *Health Technol Assess* 2003;**7**(14). <http://dx.doi.org/10.3310/hta7140>
15. Lane JA, Hamdy FC, Martin RM, Turner EL, Neal DE, Donovan JL. Latest results from the UK trials evaluating prostate cancer screening and treatment: the CAP and ProtecT studies. *Eur J Cancer* 2010;**46**:3095–101. <http://dx.doi.org/10.1016/j.ejca.2010.09.016>

16. Martin RM, Gunnell D, Hamdy F, Neal D, Lane A, Donovan J. Continuing controversy over monitoring men with localized prostate cancer: a systematic review of programs in the prostate specific antigen era. *J Urol* 2006;**176**:439–49. <http://dx.doi.org/10.1016/j.juro.2006.03.030>
17. Dall'Era MA, Albertsen PC, Bangma C, Carroll PR, Carter HB, Cooperberg MR, et al. Active surveillance for prostate cancer: a systematic review of the literature. *Eur Urol* 2012;**62**:976–83. <http://dx.doi.org/10.1016/j.eururo.2012.05.072>
18. van den Bergh RCN, Roemeling S, Roobol MJ, Wolters T, Bangma CH. Prostate-specific antigen kinetics in clinical decision-making during active surveillance for early prostate cancer – a review. *Eur Urol* 2008;**54**:505–16. <http://dx.doi.org/10.1016/j.eururo.2008.06.040>
19. Klotz L. Active surveillance for favorable risk prostate cancer: rationale, risks, and results. *Urol Oncol* 2007;**25**:505–9. <http://dx.doi.org/10.1016/j.urolonc.2007.05.021>
20. Choo R, Klotz L, Danjoux C, Morton GC, DeBoer G, Szumacher E, et al. Feasibility study: watchful waiting for localized low to intermediate grade prostate carcinoma with selective delayed intervention based on prostate specific antigen, histological and/or clinical progression. *J Urol* 2002;**167**:1664–9. [http://dx.doi.org/10.1016/S0022-5347\(05\)65174-9](http://dx.doi.org/10.1016/S0022-5347(05)65174-9)
21. Chen WM, Yang CR, Ou YC, Cheng CL, Kao YL, Ho HC, et al. Clinical outcome of patients with stage T1a prostate cancer. *J Chin Med Assoc* 2003;**66**:236–40.
22. Patel MI, DeConcini DT, Lopez-Corona E, Otori M, Wheeler T, Scardino PT. An analysis of men with clinically localized prostate cancer who deferred definitive therapy. *J Urol* 2004;**171**:1520–4. <http://dx.doi.org/10.1097/01.ju.0000118224.54949.78>
23. Mohler JL, Williams BT, Freeman JA, Mohler J. Expectant management as an option for men with stage T1c prostate cancer: a preliminary study. *World J Urol* 1997;**15**:364–8. <http://dx.doi.org/10.1007/BF01300184>
24. Khan MA, Carter HB, Epstein JI, Miller MC, Landis P, Walsh PW, et al. Can prostate specific antigen derivatives and pathological parameters predict significant change in expectant management criteria for prostate cancer? *J Urol* 2003;**170**:2274–8. <http://dx.doi.org/10.1097/01.ju.0000097124.21878.6b>
25. O'Brien MF, Cronin AM, Fearn PA, Savage CJ, Smith B, Stasi J, et al. Evaluation of prediagnostic prostate-specific antigen dynamics as predictors of death from prostate cancer in patients treated conservatively. *Int J Cancer* 2011;**128**:2373–81. <http://dx.doi.org/10.1002/ijc.25570>
26. Iremashvili V, Manoharan M, Lokeshwar SD, Rosenberg DL, Pan D, Soloway MS. Comprehensive analysis of post-diagnostic prostate-specific antigen kinetics as predictor of a prostate cancer progression in active surveillance patients. *BJU Int* 2013;**111**:396–403. <http://dx.doi.org/10.1111/j.1464-410X.2012.11295.x>
27. Vickers AJ, Savage C, O'Brien MF, Lilja H. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. *J Clin Oncol* 2009;**27**:398–403. <http://dx.doi.org/10.1200/JCO.2008.18.1685>
28. Ross AE, Loeb S, Landis P, Partin AW, Epstein JI, Kettermann A, et al. Prostate-specific antigen kinetics during follow-up are an unreliable trigger for intervention in a prostate cancer surveillance program. *J Clin Oncol* 2010;**28**:2810–16. <http://dx.doi.org/10.1200/JCO.2009.25.7311>
29. Svatek RS, Shulman M, Choudhary PK, Benaim E. Critical analysis of prostate-specific antigen doubling time calculation methodology. *Cancer* 2006;**106**:1047–53. <http://dx.doi.org/10.1002/cncr.21696>
30. Daskivich TJ, Regan MM, Oh WK. Prostate-specific antigen doubling time calculation: not as easy as 1, 2, 4. *J Urol* 2006;**176**:1927–37. <http://dx.doi.org/10.1016/j.juro.2006.07.002>

31. Connolly D, Black A, Murray LJ, Napolitano G, Gavin A, Keane PF. Methods of calculating prostate-specific antigen velocity. *Eur Urol* 2007;**52**:1044–51. <http://dx.doi.org/10.1016/j.eururo.2006.12.017>
32. Yu X, Han M, Loeb S, Gashti SN, Yeh JT, Roehl KA, *et al.* Comparison of methods for calculating prostate specific antigen velocity. *J Urol* 2006;**176**:2427–31. <http://dx.doi.org/10.1016/j.juro.2006.08.006>
33. Metcalfe C, Tilling K, Davis M, Lane J, Martin R, Kynaston H, *et al.* Current strategies for monitoring men with localised prostate cancer lack a strong evidence base: observational longitudinal study. *Br J Cancer* 2009;**101**:390–4. <http://dx.doi.org/10.1038/sj.bjc.6605181>
34. Inoue LYT, Etzioni R, Slate EH, Morrell C, Penson DF. Combining longitudinal studies of PSA. *Biostatistics*. 2004;**5**:483–500. <http://dx.doi.org/10.1093/biostatistics/kxh003>
35. Tilling K, Garmo H, Metcalfe C, Holmberg L, Hamdy FC, Neal DE, *et al.* Development of a new method for monitoring prostate-specific antigen changes in men with localised prostate cancer: a comparison of observational cohorts. *Eur Urol* 2010;**57**:446–52. <http://dx.doi.org/10.1016/j.eururo.2009.03.023>
36. Nichols JH, Loeb S, Metter EJ, Ferrucci L, Carter HB. The relationship between prostate volume and prostate-specific antigen variability: data from the Baltimore Longitudinal Study of Aging and the Johns Hopkins Active Surveillance Program. *BJU Int* 2012;**109**:1304–8. <http://dx.doi.org/10.1111/j.1464-410X.2011.10663.x>
37. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;**38**:963–74. <http://dx.doi.org/10.2307/2529876>
38. Goldstein H. Multilevel mixed linear-model analysis using iterative generalized least-squares. *Biometrika* 1986;**73**:43–56. <http://dx.doi.org/10.1093/biomet/73.1.43>
39. Davison B, Oliffe J, Pickles T, Mroz L. Factors influencing men undertaking active surveillance for the management of low-risk prostate cancer. *Oncol Nurs Forum* 2009;**36**:89–96. <http://dx.doi.org/10.1188/09.ONF.89-96>
40. Cohen H, Britten N. Who decides about prostate cancer treatment? A qualitative study. *Fam Pract* 2003;**20**:724–9. <http://dx.doi.org/10.1093/fampra/cm617>
41. Mroz LW, Oliffe JL, Davison BJ. Masculinities and patient perspectives of communication about active surveillance for prostate cancer. *Health Psychol* 2013;**32**:83–90. <http://dx.doi.org/10.1037/a0029934>
42. Volk RJ, McFall SL, Cantor SB, Byrd TL, Le YC, Kuban DA, *et al.* 'It's not like you just had a heart attack': decision-making about active surveillance by men with localized prostate cancer. *Psycho-Oncology* 2014;**23**:467–72. <http://dx.doi.org/10.1002/pon.3444>
43. Xu J, Neale AV, Dailey RK, Eggly S, Schwartz KL. Patient perspective on watchful waiting/active surveillance for localized prostate cancer. *J Am Board Fam Med* 2012;**25**:763–70. <http://dx.doi.org/10.3122/jabfm.2012.06.120128>
44. McGregor S. What information patients with localised prostate cancer hear and understand. *Patient Educ Couns* 2003;**49**:273–8. [http://dx.doi.org/10.1016/S0738-3991\(02\)00182-9](http://dx.doi.org/10.1016/S0738-3991(02)00182-9)
45. O'Rourke ME, Germino BB. Prostate cancer treatment decisions: a focus group exploration. *Oncol Nurs Forum* 1998;**25**:97–104.
46. Steginga SK, Turner E, Donovan J. The decision-related psychosocial concerns of men with localised prostate cancer: targets for intervention and research. *World J Urol* 2008;**26**:469–74. <http://dx.doi.org/10.1007/s00345-008-0279-7>

47. Sommers BD, Beard CJ, D'Amico AV, Kaplan I, Richie JP, Zeckhauser RJ. Predictors of patient preferences and treatment choices for localized prostate cancer. *Cancer* 2008;**113**:2058–67. <http://dx.doi.org/10.1002/cncr.23807>
48. Zeliadt SB, Ramsey SD, Penson DF, Hall IJ, Ekwueme DU, Stroud L, *et al.* Why do men choose one treatment over another? *Cancer* 2006;**106**:1865–74. <http://dx.doi.org/10.1002/cncr.21822>
49. Kazer MW, Bailey DE Jr, Colberg J, Kelly WK, Carroll P. The needs for men undergoing active surveillance (AS) for prostate cancer: results of a focus group study. *J Clin Nurs* 2011;**20**:581–6. <http://dx.doi.org/10.1111/j.1365-2702.2010.03489.x>
50. Davison BJ, Goldenberg SL, Gleave ME, Degner LF. Provision of individualized information to men and their partners to facilitate treatment decision making in prostate cancer. *Oncol Nurs Forum* 2003;**30**:107–14. <http://dx.doi.org/10.1188/03.ONF.107-114>
51. Davison BJ, Goldenberg SL, Wiens KP, Gleave ME. Comparing a generic and individualized information decision support intervention for men newly diagnosed with localized prostate cancer. *Cancer Nurs* 2007;**30**:E7–15. <http://dx.doi.org/10.1097/01.NCC.0000290819.22195.d6>
52. Steginga SK, Ferguson M, Clutton S, Gardiner RA, Nicol D. Early decision and psychosocial support intervention for men with localised prostate cancer: an integrated approach. *Support Care Cancer* 2008;**16**:821–9. <http://dx.doi.org/10.1007/s00520-007-0351-7>
53. Oliffe JL, Davison BJ, Pickles T, Mróz L. The self-management of uncertainty among men undertaking active surveillance for low-risk prostate cancer. *Qual Health Res* 2009;**19**:432–43. <http://dx.doi.org/10.1177/1049732309332692>
54. Hedestig O, Sandman PO, Widmark A. Living with untreated localized prostate cancer: a qualitative analysis of patient narratives. *Cancer Nurs* 2003;**26**:55–60. <http://dx.doi.org/10.1097/00002820-200302000-00008>
55. Bailey DE, Wallace M, Mishel MH. Watching, waiting and uncertainty in prostate cancer. *J Clin Nurs* 2007;**16**:734–41. <http://dx.doi.org/10.1111/j.1365-2702.2005.01545.x>
56. van den Bergh RC, van Vugt HA, Korfage IJ, Steyerberg EW, Roobol MJ, Schroder FH, *et al.* Disease insight and treatment perception of men on active surveillance for early prostate cancer. *BJU Int* 2010;**105**:322–8. <http://dx.doi.org/10.1111/j.1464-410X.2009.08764.x>
57. Pickles T, Ruether JD, Weir L, Carlson L, Jakulj F. Psychosocial barriers to active surveillance for the management of early prostate cancer and a strategy for increased acceptance. *BJU Int* 2007;**100**:544–51. <http://dx.doi.org/10.1111/j.1464-410X.2007.06981.x>
58. van den Bergh RC, Korfage IJ, Bangma CH. Psychological aspects of active surveillance. *Curr Opin Urol* May 2012;**22**:237–42. <http://dx.doi.org/10.1097/MOU.0b013e328351dcb1>
59. van den Bergh RCN, Essink-Bot M-L, Roobol MJ, Wolters T, Schröder FH, Bangma CH, *et al.* Anxiety and distress during active surveillance for early prostate cancer. *Cancer* 2009;**115**:3868–78. <http://dx.doi.org/10.1002/cncr.24446>
60. Burnet KL, Parker C, Dearnaley D, Brewin CR, Watson M. Does active surveillance for men with localized prostate cancer carry psychological morbidity? *BJU Int* 2007;**100**:540–3. <http://dx.doi.org/10.1111/j.1464-410X.2007.07009.x>
61. Seiler D, Randazzo M, Leupold U, Zeh N, Isbarn H, Chun FK, *et al.* Protocol-based active surveillance for low-risk prostate cancer: anxiety levels in both men and their partners. *Urology* 2012;**80**:564–9. <http://dx.doi.org/10.1016/j.urology.2012.04.053>
62. Vasarainen H, Lokman U, Ruutu M, Taari K, Rannikko A. Prostate cancer active surveillance and health-related quality of life: results of the Finnish arm of the prospective trial. *BJU Int* 2012;**109**:1614–19. <http://dx.doi.org/10.1111/j.1464-410X.2011.10677.x>

63. van den Bergh RCN, Essink-Bot M-L, Roobol MJ, Schröder FH, Bangma CH, Steyerberg EW. Do anxiety and distress increase during active surveillance for low risk prostate cancer? *J Urol* 2010;**183**:1786–91. <http://dx.doi.org/10.1016/j.juro.2009.12.099>
64. Bellardita L, Rancati T, Alvisi MF, Villani D, Magnani T, Marengi C, *et al.* Predictors of health-related quality of life and adjustment to prostate cancer during active surveillance. *Eur Urol* 2013;**64**:30–6. <http://dx.doi.org/10.1016/j.eururo.2013.01.009>
65. Dale W, Bilir P, Han M, Meltzer D. The role of anxiety in prostate carcinoma: a structured review of the literature. *Cancer* 2005;**104**:467–78. <http://dx.doi.org/10.1002/cncr.21198>
66. Bergman J, Litwin MS. Quality of life in men undergoing active surveillance for localized prostate cancer. *J Nat Cancer Inst Monogr* 2012;**2012**:242–9. <http://dx.doi.org/10.1093/jncimonographs/lgs026>
67. Couper JW, Love AW, Dunai JV, Duchesne GM, Bloch S, Costello AJ, *et al.* The psychological aftermath of prostate cancer treatment choices: a comparison of depression, anxiety and quality of life outcomes over the 12 months following diagnosis. *Med J Aust* 2009;**190**(Suppl. 7):86–9.
68. National Institute for Health and Care Excellence. *Prostate Cancer: Diagnosis and Treatment*. NICE Clinical Guideline 175; 2014.
69. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58. <http://dx.doi.org/10.1002/sim.1186>
70. StataCorp. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP; 2011.
71. Allué López M, Allepuz Losa C, Borque Fernando A, Serrano Frago P, Gil Martínez P, Gil Sanz M. Cáncer de próstata incidental: T1a-T1b. Nuestra experiencia tras observación/intervención radical y revisión de la literatura. *Actas Urol Esp* 2006;**30**:749–53. [http://dx.doi.org/10.1016/S0210-4806\(06\)73531-8](http://dx.doi.org/10.1016/S0210-4806(06)73531-8)
72. Al Otaibi M, Ross P, Fahmy N, Jeyaganth S, Trottier H, Sircar K, *et al.* Role of repeated biopsy of the prostate in predicting disease progression in patients with prostate cancer on active surveillance. *Cancer* 2008;**113**:286–92. <http://dx.doi.org/10.1002/cncr.23575>
73. Descazeaud A, Peyromaure M, Salin A, Amsellem-Ouazana D, Flam T, Viellefond A, *et al.* Predictive factors for progression in patients with clinical stage T1a prostate cancer in the PSA era. *Eur Urol* 2008;**53**:355–62. <http://dx.doi.org/10.1016/j.eururo.2007.06.020>
74. Ercole B, Marietti SR, Fine J, Albertsen PC. Outcomes following active surveillance of men with localized prostate cancer diagnosed in the prostate-specific antigen era. *J Urol* 2008;**180**:1336–41. <http://dx.doi.org/10.1016/j.juro.2008.06.027>
75. Roemeling S, Roobol MJ, de Vries SH, Wolters T, Gosselaar C, van Leenders GJLH, *et al.* Active surveillance for prostate cancers detected in three subsequent rounds of a screening trial: characteristics, PSA doubling times, and outcome. *Eur Urol* 2007;**51**:1244–51. <http://dx.doi.org/10.1016/j.eururo.2006.11.053>
76. Hardie C, Parker C, Norman A, Eeles R, Horwich A, Huddart R, *et al.* Early outcomes of active surveillance for localized prostate cancer. *BJU Int* 2005;**95**:956–60. <http://dx.doi.org/10.1111/j.1464-410X.2005.05446.x>
77. Adamy A, Yee DS, Matsushita K, Maschino A, Cronin A, Vickers A, *et al.* Role of prostate specific antigen and immediate confirmatory biopsy in predicting progression during active surveillance for low risk prostate cancer. *J Urol* 2011;**185**:477–82. <http://dx.doi.org/10.1016/j.juro.2010.09.095>
78. Soloway MS, Soloway CT, Eldefrawy A, Acosta K, Kava B, Manoharan M. Careful selection and close monitoring of low-risk prostate cancer patients on active surveillance minimizes the need for treatment. *Eur Urol* 2010;**58**:831–5. <http://dx.doi.org/10.1016/j.eururo.2010.08.027>

79. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol* 2010;**28**:126–31. <http://dx.doi.org/10.1200/JCO.2009.24.2180>
80. Godtman RA, Holmberg E, Khatami A, Stranne J, Hugosson K. Outcome following surveillance of men with screen-detected prostate cancer. Results from the Gothenburg randomised population-based prostate cancer screening trial. *Eur Urol Suppl* 2012;**11**:e1094–e1094a. [http://dx.doi.org/10.1016/S1569-9056\(12\)61090-2](http://dx.doi.org/10.1016/S1569-9056(12)61090-2)
81. Cooperberg MR, Cowan JE, Hilton JF, Reese AC, Zaid HB, Porten SP, *et al.* Outcomes of active surveillance for men with intermediate-risk prostate cancer. *J Clin Oncol* 2011;**29**:228. <http://dx.doi.org/10.1200/JCO.2010.31.4252>
82. Tosoian JJ, Trock BJ, Landis P, Feng Z, Epstein JI, Partin AW, *et al.* Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *J Clin Oncol* 2011;**29**:2185. <http://dx.doi.org/10.1200/JCO.2010.32.8112>
83. San Francisco IF, Werner L, Regan MM, Garnick MB, Bubley G, DeWolf WC. Risk stratification and validation of prostate specific antigen density as independent predictor of progression in men with low risk prostate cancer during active surveillance. *J Urol* 2011;**185**:471–6. <http://dx.doi.org/10.1016/j.juro.2010.09.115>
84. Kakehi Y, Kamoto T, Shiraishi T, Ogawa O, Suzukamo Y, Fukuhara S, *et al.* Prospective evaluation of selection criteria for active surveillance in Japanese patients with stage T1cN0M0 prostate cancer. *Jpn J Clin Oncol* 2008;**38**:122–8. <http://dx.doi.org/10.1093/jjco/hym161>
85. Selvadurai ED, Singhera M, Thomas K, Mohammed K, Woode-Amisshah R, Horwich A, *et al.* Medium-term outcomes of active surveillance for localised prostate cancer. *Eur Urol* 2013;**64**:981–7. <http://dx.doi.org/10.1016/j.eururo.2013.02.020>
86. Ischia JJ, Pang CY, Tay YK, Li Wai Suen CFD, Aw HC, Frydenberg M. Active surveillance for prostate cancer: an Australian experience. *BJU Int* 2012;**109**:40–3. <http://dx.doi.org/10.1111/j.1464-410X.2012.11045.x>
87. Miocinovic R, Jones JS, Pujara AC, Klein EA, Stephenson AJ. Acceptance and durability of surveillance as a management choice in men with screen-detected, low-risk prostate cancer: improved outcomes with stringent enrollment criteria. *Urology* 2011;**77**:980–4. <http://dx.doi.org/10.1016/j.urology.2010.09.063>
88. Ayres BE, Montgomery BSI, Barber NJ, Pereira N, Langley SE, Denham P, *et al.* The role of transperineal template prostate biopsies in restaging men with prostate cancer managed by active surveillance. *BJU Int* 2012;**109**:1170–6. <http://dx.doi.org/10.1111/j.1464-410X.2011.10480.x>
89. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, *et al.* Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. *Eur Urol* 2012;**63**:597–603. <http://dx.doi.org/10.1016/j.eururo.2012.11.005>
90. Cookson MS, Fleshner NE, Soloway SM, Fair WR. Correlation between Gleason score of needle biopsy and radical prostatectomy specimen: accuracy and clinical implications. *J Urol* 1997;**157**:559–62. [http://dx.doi.org/10.1016/S0022-5347\(01\)65201-7](http://dx.doi.org/10.1016/S0022-5347(01)65201-7)
91. Lattouf JB, Saad F. Gleason score on biopsy: is it reliable for predicting the final grade on pathology? *BJU Int* 2002;**90**:694–8. <http://dx.doi.org/10.1046/j.1464-410X.2002.02990.x>
92. Grossfeld GD, Chang JJ, Broering JM, Li YUP, Lubeck DP, Flanders SC, *et al.* Under staging and under grading in a contemporary series of patients undergoing radical prostatectomy: results from the Cancer of the Prostate Strategic Urologic Research Endeavor database. *J Urol* 2001;**165**:851–6. [http://dx.doi.org/10.1016/S0022-5347\(05\)66543-3](http://dx.doi.org/10.1016/S0022-5347(05)66543-3)

93. Simpkin AJ, Metcalfe C, Martin RM, Lane JA, Donovan JL, Hamdy FC, *et al.* Longitudinal prostate-specific antigen reference ranges: choosing the underlying model of age-related changes. *Stat Methods Med Res* 2015; in press.
94. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep.* 1966;**50**:125.
95. Patel A, Dorey F, Franklin J, deKernion JB. Recurrence patterns after radical retropubic prostatectomy: clinical usefulness of prostate specific antigen doubling times and log slope prostate specific antigen. *J Urol* 1997;**158**:1441–5. [http://dx.doi.org/10.1016/S0022-5347\(01\)64238-1](http://dx.doi.org/10.1016/S0022-5347(01)64238-1)
96. Zhang L, Loblaw A, Klotz L. Modeling prostate specific antigen kinetics in patients on active surveillance. *J Urol* 2006;**176**:1392–8. <http://dx.doi.org/10.1016/j.juro.2006.06.103>
97. Pan H, Goldstein H. Multi-level models for longitudinal growth norms. *Stat Med* 1997;**16**:2665–78. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19971215\)16:23<2665::AID-SIM711>3.0.CO;2-V](http://dx.doi.org/10.1002/(SICI)1097-0258(19971215)16:23<2665::AID-SIM711>3.0.CO;2-V)
98. Tilling K, Sterne JAC, Wolfe CDA. Multilevel growth curve models with covariate effects: application to recovery after stroke. *Stat Med* 2001;**20**:685–704. <http://dx.doi.org/10.1002/sim.697>
99. MATLAB. *Version 7.14 (R2012a)*. Natick, MA: The MathWorks Inc.; 2012.
100. Macdonald-Wallis C, Lawlor DA, Palmer T, Tilling K. Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy. *Stat Med* 2012;**31**:3147–64. <http://dx.doi.org/10.1002/sim.5385>
101. van As NJ, Norman AR, Thomas K, Khoo VS, Thompson A, Huddart RA, *et al.* Predicting the probability of deferred radical treatment for localised prostate cancer managed by active surveillance. *Eur Urol* 2008;**54**:1297–1305. <http://dx.doi.org/10.1016/j.eururo.2008.02.039>
102. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* 2005;**293**:2095–101. <http://dx.doi.org/10.1001/jama.293.17.2095>
103. Loeb S, Kettermann A, Ferrucci L, Landis P, Metter EJ, Carter HB. PSA doubling time versus PSA velocity to predict high-risk prostate cancer: data from the Baltimore Longitudinal Study of Aging. *Eur Urol* 2008;**54**:1073–80. <http://dx.doi.org/10.1016/j.eururo.2008.06.076>
104. Blanker M, Groeneveld F, Prins A, Bernsen R, Bohnen A, Bosch J. Strong effects of definition and nonresponse bias on prevalence rates of clinical benign prostatic hyperplasia: the Krimpen study of male urogenital tract problems and general health status. *BJU Int* 2000;**85**:665–71. <http://dx.doi.org/10.1046/j.1464-410x.2000.00570.x>
105. Cox DR. Regression models and life tables. *J R Stat Soc Series B Methodol* 1972;**34**:187–220.
106. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;**247**:2543–6. <http://dx.doi.org/10.1001/jama.1982.03320430047030>
107. Glaser BG, Strauss AL. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine de Gruyter; 1967.
108. Parulekar WR, McKenzie M, Chi KN, Klotz C, Catton C, Brundage M, *et al.* Defining the optimal treatment strategy for localized prostate cancer patients: a survey of ongoing studies at the National Cancer Institute of Canada Clinical Trials Group. *Current Oncol* 2008;**15**:179. <http://dx.doi.org/10.3747/co.v15i4.257>
109. Donovan J, Mills N, Smith M, *et al.* Quality improvement report: improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *BMJ* 2002;**325**:766. <http://dx.doi.org/10.1136/bmj.325.7367.766>

110. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, *et al.* Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;**280**:969–74. <http://dx.doi.org/10.1001/jama.280.11.969>
111. Hernandez DJ, Nielsen ME, Han M, Partin AW. Contemporary evaluation of the D'Amico risk classification of prostate cancer. *Urology* 2007;**70**:931–5. <http://dx.doi.org/10.1016/j.urology.2007.08.055>
112. Vickers AJ, Ulmert D, Sjöberg DD, Bennette CJ, T Björk, Gerdtsson A, *et al.* Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. *BMJ* 2013;**346**:f2023. <http://dx.doi.org/10.1136/bmj.f2023>
113. Coakley FV, Qayyum A, Kurhanewicz J. Magnetic resonance imaging and spectroscopic imaging of prostate cancer. *J Urol* 2003;**170**:S69–76. <http://dx.doi.org/10.1097/01.ju.0000094958.23276.c4>
114. Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur S, *et al.* TMPRSS2: ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* 2007;**26**:4596–9. <http://dx.doi.org/10.1038/sj.onc.1210237>
115. Shaw G, Thomas B, Dawson S, Srivastava G, Vowler S, Gnanapragasam V, *et al.* Identification of pathologically insignificant prostate cancer is not accurate in unscreened men. *Br J Cancer* 2014;**110**:2405–11. <http://dx.doi.org/10.1038/bjc.2014.192>
116. Sölétormos G, Semjonow A, Sibley PEC, Lamerz R, Petersen PH, Albrecht W, *et al.* Biological variation of total prostate-specific antigen: a survey of published estimates and consequences for clinical practice. *Clin Chem* 2005;**51**:1342–51. <http://dx.doi.org/10.1373/clinchem.2004.046086>
117. Makarov DV, Isharwal S, Sokoll LJ, Landis P, Marlow C, Epstein JI, *et al.* Pro–prostate-specific antigen measurements in serum and tissue are associated with treatment necessity among men enrolled in expectant management for prostate cancer. *Clin Cancer Res* 2009;**15**:7316. <http://dx.doi.org/10.1158/1078-0432.CCR-09-1263>
118. Haese A, de la Taille A, Van Poppel H, Marberger M, Stenzl A, Mulders PFA, *et al.* Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. *Eur Urol* 2008;**54**:1081–8. <http://dx.doi.org/10.1016/j.eururo.2008.06.071>
119. Bastian PJ, Palapattu GS, Lin X, Yegnasubramanian S, Mangold LA, Trock B, *et al.* Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. *Clin Cancer Res* 2005;**11**:4037–43. <http://dx.doi.org/10.1158/1078-0432.CCR-04-2446>
120. Coakley FV, Chen I, Qayyum A, Westphalen AC, Carroll PR, Hricak H, *et al.* Validity of prostate-specific antigen as a tumour marker in men with prostate cancer managed by watchful-waiting: correlation with findings at serial endorectal magnetic resonance imaging and spectroscopic imaging. *BJU Int* 2007;**99**:41–5. <http://dx.doi.org/10.1111/j.1464-410X.2006.06515.x>
121. van den Bergh RC, Ahmed HU, Bangma CH, Cooperberg MR, Villers A, Parker CC. Novel tools to improve patient selection and monitoring on active surveillance for low-risk prostate cancer: a systematic review. *Eur Urol* 2014;**65**:1023–31. <http://dx.doi.org/10.1016/j.eururo.2014.01.027>
122. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;**19**:349–57. <http://dx.doi.org/10.1093/intqhc/mzm042>
123. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer-Verlag; 2009.

124. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates – parsimonious parametric modelling. *Appl Stat J R Stat Soc C* 1994;**43**:429–67. <http://dx.doi.org/10.2307/2986270>
125. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Vol. 790. Chichester: John Wiley & Sons; 2008. <http://dx.doi.org/10.1002/9780470770771>
126. Long J, Ryoo J. Using fractional polynomials to model non-linear trends in longitudinal data. *Br J Math Stat Psychol* 2010;**63**:177–203. <http://dx.doi.org/10.1348/000711009X431509>
127. Mackenzie ML, Donovan C, McArdle B. Regression spline mixed models: a forestry example. *J Agric Biol Environ Stat* 2005;**10**:394–410. <http://dx.doi.org/10.1198/1085711105X80194>
128. Bellera CA, Hanley JA, Joseph L, Albertsen PC. Hierarchical change point models for biochemical markers illustrated by tracking postradiotherapy prostate-specific antigen series in men with prostate cancer. *Ann Epidemiol* 2008;**18**:270–82. <http://dx.doi.org/10.1016/j.annepidem.2007.10.006>
129. McCarthy A, Hughes R, Tilling K, Davies D, Smith GD, Ben-Shlomo Y. Birth weight; postnatal, infant, and childhood growth; and obesity in young adulthood: evidence from the Barry Caerphilly Growth Study. *Am J Clin Nutr* 2007;**86**:907–13.
130. Ramsay JO, Silverman BW. *Functional Data Analysis*. New York, NY: Springer-Verlag; 2005. <http://dx.doi.org/10.1002/0470013192.bsa239>
131. Yao F, Muller HG, Wang JL. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 2005;**100**:577–90. <http://dx.doi.org/10.1198/016214504000001745>

Appendix 1 Development of active monitoring system

The model developed in *Chapter 3* was included in an AMS through the MS Excel software, such that serial PSAs could be entered and displayed for men following an AM/AS programme. Interviews with men on AM/AS, urologists and oncologists included a demonstration of the system.

The main interface allows entry of date of birth, Gleason score, PSA test date and result. The Excel sheet then calculates one of a number of options. The first is the average trend of PSA for a similar man (i.e. with the same PSA, age and Gleason score at diagnosis) with clinically localised prostate cancer, as shown in the green line in *Figure 10*. There is also the option to show the average trend along with 95% reference range for a similar man without PCa, as shown in the blue and orange lines of *Figure 10*. Using a simple interface, men can compare their own PSA results with the average PSA growth of similar men with and without prostate cancer.

Initially, we intended to use the 95% reference range for men without PCa as the method that would alert men with PCa to rapidly rising PSA. After discussion with the steering group, the use of reference ranges of men with PCa were taken forward for use in the comparison of *Chapter 5*.

Note that data presented are simulated to be indicative of men with PCa. The names given are fictitious and are not representative of any men in any of the cohorts.

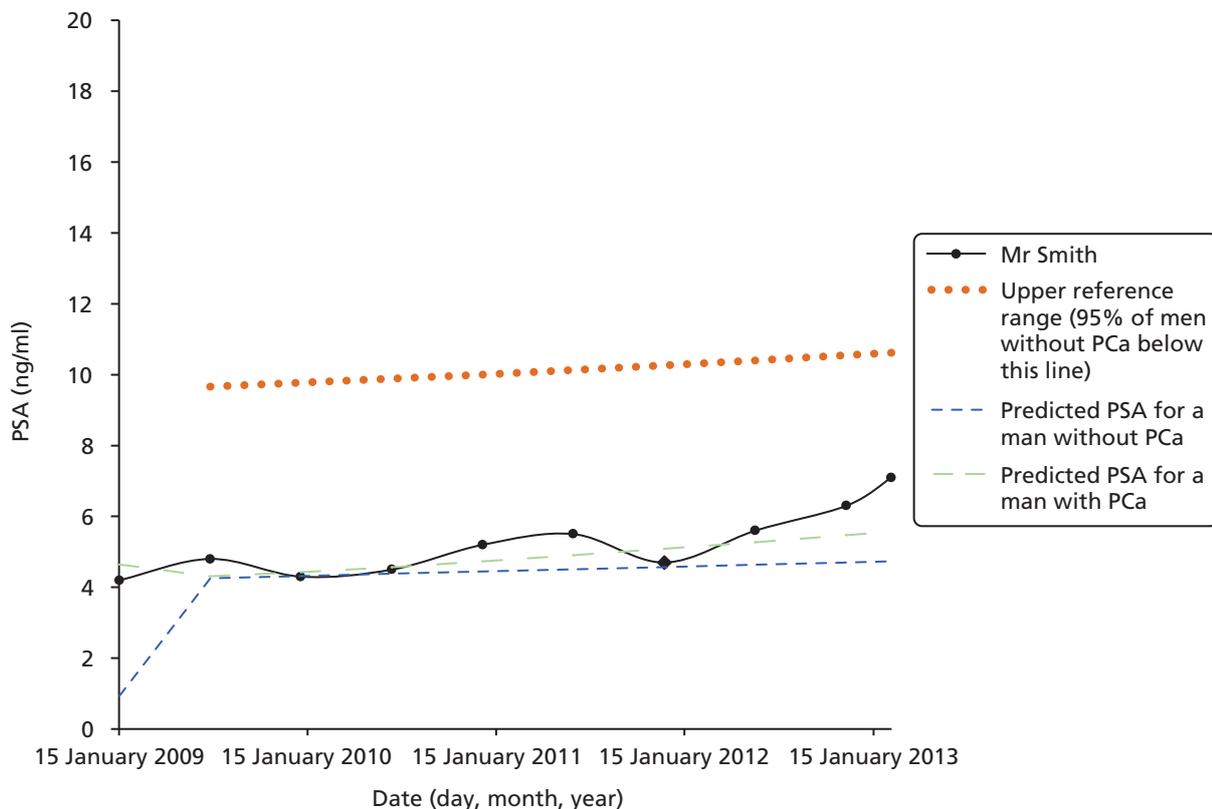


FIGURE 10 The PSARR system for a simulated individual, showing average PSA trend (blue, short dash) plus 95% reference range (orange) for men without PCa, as well as average PSA trend for men with PCa (green, long dash).

Appendix 2 Dissemination

Peer-reviewed articles

Simpkin AJ, Metcalfe C, Martin RM, Lane JA, Donovan JL, Hamdy FC, *et al.* Longitudinal prostate-specific antigen reference ranges: choosing the underlying model of age-related changes [published online ahead of print October 9 2013]. *Stat Methods Med Res* 2013.

Simpkin AJ, Tilling K, Martin RM, Lane JA, Hamdy FC, Holmberg L, *et al.* Over or under treatment? A systematic review and meta-analysis of factors determining change of management in Active Surveillance for localised prostate cancer. *Eur Urol* 2015;**67**:1006–8.

Conference oral presentations

Simpkin A. A systematic review of active monitoring for prostate cancer. Prostate Cancer Institute, Galway, January 2012.

Simpkin A. Longitudinal prostate-specific antigen reference ranges: choosing the underlying model of age-related changes. Conference on Applied Statistics in Ireland (CASI), Kildare, May 2013.

Simpkin A. Longitudinal prostate-specific antigen reference ranges: choosing the underlying model of age-related changes. International Society for Clinical Biostatistics, Munich, August 2013.

Simpkin A. A systematic review of active monitoring for prostate cancer. National Cancer Research Institute (NCRI), Liverpool, October 2013.

Conference poster presentations

Simpkin AJ, Metcalfe C, Martin RM, Lane JA, Donovan JL, Hamdy FC, *et al.* Modelling prostate-specific antigen change in men with clinically localised prostate cancer follow by active monitoring. CASI 2012, Dundalk, May 2012.

Simpkin AJ, Tilling K, Martin RM, Lane JA, Hamdy FC, Holmberg L, *et al.* Over or under treatment? A systematic review and meta-analysis of factors determining change of management in Active Surveillance for localised prostate cancer. NCRI 2013, Liverpool, October 2013.

Appendix 3 Qualitative documentation

Clinicians' interview topic guide

BOX 2 Clinicians' interview topic guide

Introduction

Researcher explains research, asks if any queries on participant information sheet (PIS), and takes written consent.

Background/context

The NICE guidelines indicate that AM/AS are options, along with radical prostatectomy and radical radiotherapy, for PCa but methods for doing this have not been clearly specified. We are, therefore, very interested to know what methods you are using.

Definitions/terminology

What terms do you use to describe men whose prostate cancer is being monitored? Do you differentiate between different subgroups among these men?

Three terms are used to describe monitoring – what do you use these terms to mean? AM? AS? WW?

Current practice

What factors would influence you to recommend AM/AS to a man, newly diagnosed with PCa?

Prompts: Age, Gleason score, PSA, % positive cores, prostate volume, comorbidities, other?

What factors would influence you to recommend a more active treatment?

Prompts: Age, Gleason score, PSA, % positive cores, prostate volume, comorbidities, other?

What do you say to men, when giving treatment options for those newly diagnosed with localised PCa, Gleason score of 6, PSA < 10 ng/ml and < 50% positive cores?

Prompts: What other factors might influence your approach? Age? Prostate volume? Comorbidities?

Routine appointments for men on AM/AS.

What do you recommend to patients regarding frequency of PSA tests? *Why?*

Factors that influence you to vary that frequency?

What is your routine for appointments where PSA findings for men on AM/AS/WW are discussed?

How do you deal with rises in PSA values? How are these presented to men? What phrases do you use?

BOX 2 Clinicians' interview topic guide (*continued*)

What helps to reassure men? What makes them more anxious?

How often will you do a rectal examination?

Factors influencing you to vary that frequency?

What are your views on rebiopsy?

Factors influencing you to recommend rebiopsy? Avoid rebiopsy?

Has your practice in how you monitor men on AM/AS/WW changed over time?

Factors influencing these changes?

How do you think your practice compares to that of your colleagues? *Same? Different – how?*

If you were asked to identify one key aspect of current practice for monitoring PSA levels that you believe could be improved, what would that be?

Demonstration of AMS

Presentation of three hypothetical patient vignettes:

Information presented in traditional, descriptive format of a case summary (see vignettes).

What would you recommend to each of these men regarding ongoing management of prostate cancer? *Rationale?*

What triggers would alert you to possible progression of the disease? What would trigger you to carry out further investigations and what would those be?

Prompts: [PSA kinetics of some sort, rebiopsy, imaging MRI or transrectal ultrasound scan (TRUS), bone scan?]

For each case use prompts as follows if not initiated by clinician:

This man had a rebiopsy after 2 years and it:

showed no disease progression – how would that influence your recommendations?

showed disease progression to Gleason score of 7 – how would that influence your recommendations?

How would your recommendations be affected if this man were 10 years older? 10 years younger? Why?

Information presented using the AMS.

What would you recommend to each of these men regarding ongoing management of prostate cancer?

BOX 2 Clinicians' interview topic guide (*continued*)*Rationale?*

What are your thoughts on how the information is presented on screen?

Probe: aspects you really like, aspects you do not like, aspects that are clear, aspects that are not clear.

What are your thoughts on the information that the system gives?

Probe: useful, less useful, missing, would like to see.

What extra information does the system give men as compared with what usually happens in clinic?

What disadvantages does the system bring compared with what usually happens in clinic?

If you were this man's clinician would the system make any difference to your thinking about the next step in treatment?

Would you use the system in its current format? If not, what changes would need to be made to make it useful to you?

What do you think would be a good name for the system?

Future trial to evaluate effectiveness and cost-effectiveness of the AMS

We would like to investigate if using the AMS is more effective for men on AM than current practice. To do so, we would run a study in which all those participating are randomly allocated to either current practice or using the AMS. What are your thoughts on such a study? Can you envisage any problems?

Wind-down and explanation of continuing research process

Researcher checks topic guide for omissions and ask whether or not any questions have arisen for the interviewee. Researcher then explains the next stages of the research process (audio recording of clinics) and assesses whether the informant is prepared to grant their ongoing consent for participation in the research.

[Offer summary of results and stop recorder.]

Men's interview topic guide

BOX 3 Men's interview topic guide

Introduction

Researcher explains research, asks if any queries on PIS, and takes consent.

Context/background

Can you tell me a bit about when you were first diagnosed with PCa and your reaction at the time?

Probe: When did they find out they had PCa? Why did they consult? What was their reaction to the diagnosis?

How long you have been on AM/AS?

How has that process been?

Probe: Frequency of blood tests, feelings around time of blood test, high points and low points.

Regular consultations to discuss PSA findings

Can you give me an overview of your current routine for having your PSA levels monitored? See doctor/nurse? Attend alone? Types and frequency of tests/examinations/biopsies undertaken.

What are your thoughts about the current approach to monitoring PSA levels?

How often do you have a blood test at present?

Any recent changes? What difference have these made?

What happens at the appointments where you discuss the result?

Probe: how are results presented? What is most helpful/least helpful about these discussions/appointments?

Can you think of something that would make these discussions easier?

Probe: Face to face, phone, length of appointment, content discussed.

How are you feeling about your next outpatient appointment?

Probe: are they anxious/neutral/relaxed; what issues cause concern.

What sorts of things does your doctor/nurse say to you when your PSA rises?

What effect does this have on how you feel?

BOX 3 Men's interview topic guide (*continued*)**Demonstration of AMS**

Presentation:

What are your thoughts on how the information is presented on screen?

Probe: aspects you really like, aspects you do not like, aspects that are clear, aspects that are not clear.

Information:

What are your thoughts on the information that the system gives?

Probe: useful, less useful, missing, would like to see.

Presentation of three case-study vignettes using traditional case description (see case vignettes) of information that would be available in clinic routinely under current practice.

Probe for each vignette: if you were this man and had this information what would your thinking be? Why?

Presentation of three case study vignettes using AMS.

Probes: if you were this man and had this information what would your thinking be? Why?

How does the information given by the AMS compare to the information you usually have in clinic?

What disadvantages does the system bring compared with what usually happens in clinic?

If you were in this man's position would the system make any difference to your thinking about the next step for you?

Would you use this system as it currently is? If not, what would have to change for you to want to use it?

What do you think would be a good name for the system?

Future trial to evaluate effectiveness and cost-effectiveness of the AMS

We would like to investigate whether using the AMS is more effective for men on AM than current practice. To do so, we would run a study in which all those participating are allocated to either current practice or using the AMS. What do you think about taking part in such a study?

Allocation to current practice or AMS will be by chance, so you have an equal chance of getting either. What do you think about taking part in such a study?

Wind-down and explanation of continuing research process

Researcher checks topic guide for omissions and asks whether any questions have arisen that remain unanswered.

Researcher then explains the next stages of the research process (audio-recording of clinic appointment) and assesses whether or not the informant is prepared to grant their ongoing consent for participation in the research. Repeat measure of anxiety. [Offer summary of results and stop recorder.]

Paper vignettes

TABLE 20 Mr Smith paper vignette

Date	Time on monitoring (months)	PSA (ng/ml)	PSADT (years)	PSAv (ng/ml/year)
15 January 2009	0	4.2		
10 July 2009	6	4.8		
2 January 2010	12	4.3	19.24	0.04
27 June 2010	18	4.5	27.94	0.02
20 December 2010	24	5.2	8.66	0.08
14 June 2011	30	5.5	6.94	0.10
7 December 2011	36	4.7	11.59	0.06
31 May 2012	42	5.6	9.93	0.07
23 November 2012	48	6.3	8.04	0.09
18 February 2013	54	7.1	6.63	0.10

TABLE 21 Mr Jones paper vignette

Date	Time on monitoring (months)	PSA (ng/ml)	PSADT (years)	PSAv (ng/ml/year)
9 February 2010	0	3.1		
10 July 2010	5	7.5		
8 December 2010	10	2.1	-1.42	-0.49
8 May 2011	15	8.7	1.50	0.46
6 October 2011	20	5.4	2.22	0.31
5 March 2012	25	9.8	1.62	0.43
3 August 2012	30	3.2	4.79	0.14
1 January 2013	35	7.6	3.80	0.18

TABLE 22 Mr Evans paper vignette

Date	Time on monitoring (months)	PSA (ng/ml)	PSADT (years)	PSAv (ng/ml/year)
3 August 2004	0	4.2		
19 February 2005	6	4.6		
7 September 2005	12	3.8	-7.04	-0.098
26 March 2006	18	4.5	-476.9	-0.0014
12 October 2006	24	5.1	10.57	0.065
30 April 2007	30	5.3	8.14	0.085
16 November 2007	36	3.9	31.82	0.021
3 June 2008	42	4.8	26.76	0.025
20 December 2008	48	6.4	11.33	0.061
8 July 2009	54	4.1	21.82	0.031
24 January 2010	60	6.7	12.66	0.054
12 August 2010	66	7.2	9.78	0.071
28 February 2011	72	7.8	8.34	0.083
16 September 2011	78	8.5	7.46	0.092
3 April 2012	84	7.8	7.46	0.092
20 October 2012	90	10.2	6.81	0.101

Active monitoring system vignettes

Mr Smith active monitoring system vignette

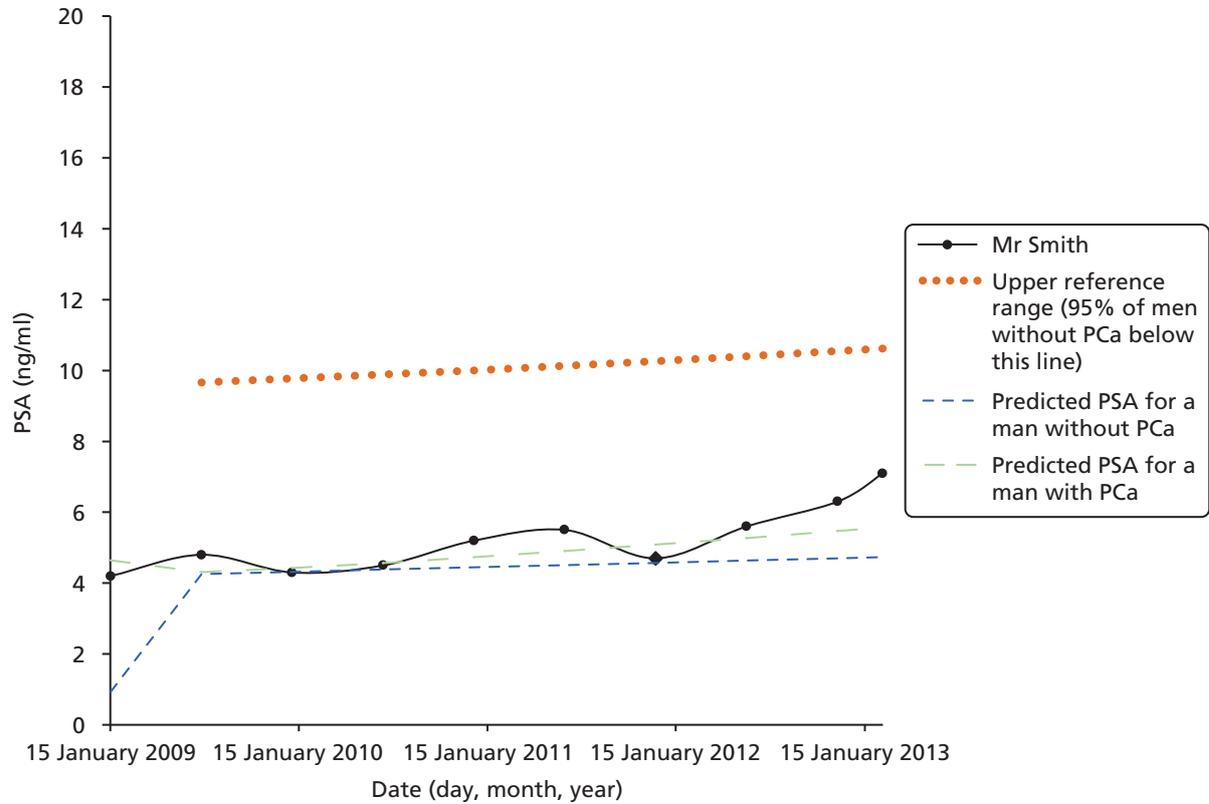
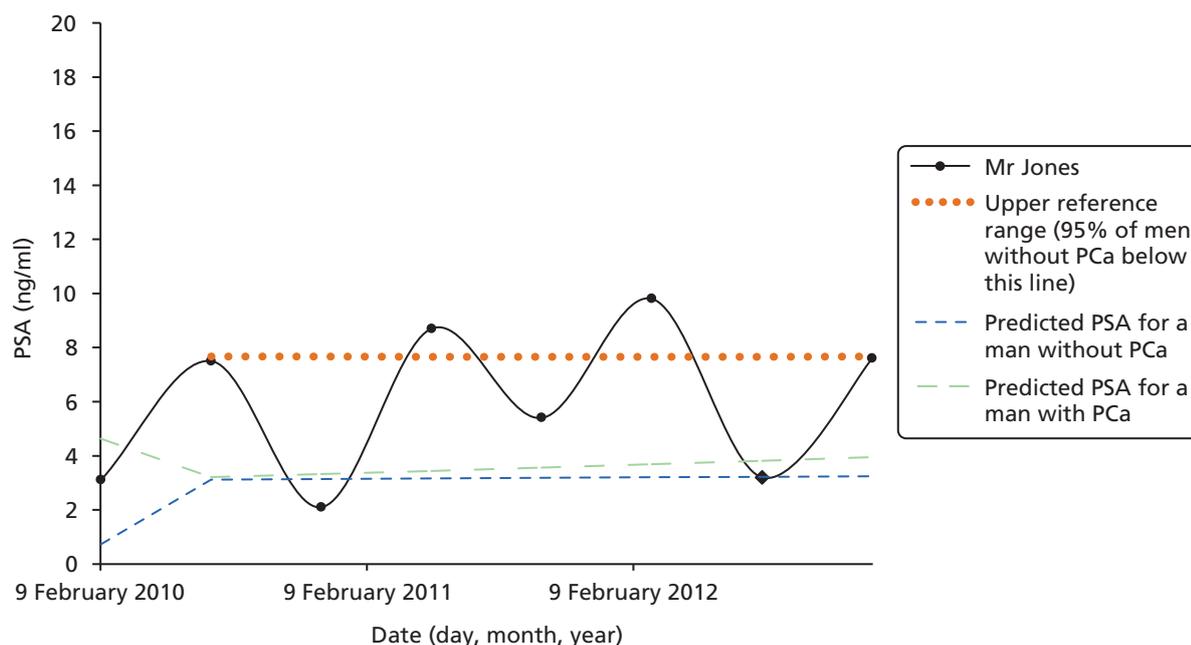
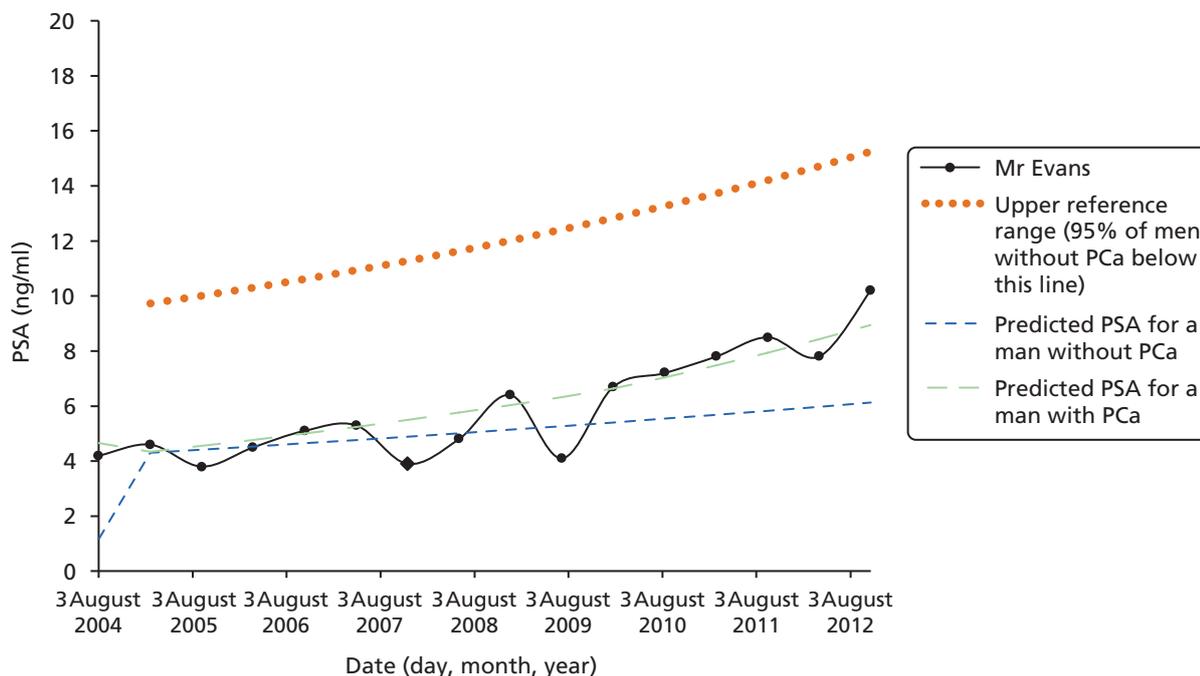


FIGURE 11 Mr Smith vignette presented on AMS.

Mr Jones active monitoring system vignette**FIGURE 12** Mr Jones vignette presented on AMS.**Mr Evans active monitoring system vignette****FIGURE 13** Mr Evans vignette presented on AMS.

Clinicians' definitions of 'active surveillance' versus 'watchful waiting'

Clinicians were asked what they understood about the term 'AS', and how they distinguished this from 'WW' – another term used by clinicians across all sites. There was broad agreement over what AS and WW programmes aimed to do, how they achieved this and whom they targeted.

Most participants touched on the idea that AS had a curative intent, whereas WW was more aligned with symptom management and palliative care:

The point of active surveillance is that you have the intention to treat or to cure the patient. [Later] Watchful waiting is very much, it's not a curative thing – it's there just to see as symptoms come up, we try and treat those symptoms.

U14

Some clinicians talked about differences in the types of treatments offered within each programme. Whereas AS patients were eligible to receive radical treatments (i.e. surgery and radiotherapy), patients on WW received symptomatic relief (e.g. hormone treatment):

Well we usually use active surveillance or watchful waiting, where active surveillance means they're under surveillance but there's still a potential that they could have radical treatment if the parameters change, whereas watchful waiting is for the patient who would then have hormone therapy if their parameters change.

U2

One clinician expressed a slightly different view to others, in that he identified two groups within the AS bracket: one group for whom both surgery and radiotherapy would be suitable and another group for whom radiotherapy alone would be suitable:

[. . .] the classic active monitoring or active surveillance are, I think for patients with low risk prostate cancer so low Gleason score, low volume, low PSA, who often are fairly young, you know certainly below 70 and if their disease progresses you would be considering probably a radical prostatectomy or one of the other radical treatments. There's also a group of patients that we call active surveillance who are a little bit more, a little bit different to that, who are the patients who are probably a bit older, may have slightly more aggressive disease but still fairly low risk and who, if they progressed you would be considering radical radiotherapy as in they wouldn't be suitable for surgery, but radiotherapy.

U5

One distinction that arose was the tendency to associate AS with younger and WW with more elderly patients. One clinician suggested that the division was not about age, but more to do with physical status. Men on AS needed to be fit enough to undergo radical therapies, whereas men on WW were deemed unsuitable for this:

Watchful waiting is very much for, erm, men who are not medically fit, basically. It doesn't necessarily mean that age – it does figure into it but if you're 85 you're not going to benefit from surgery [. . .]. At the end of the day the point of active surveillance is that you have the intention to treat or to cure the patient, so they have to be medically fit enough to undergo that procedure itself.

U14

Clinicians' views on active surveillance eligibility

Gleason score

There was little variation within or across sites when it came to clinicians' interpretations of 'appropriate' Gleason scores for AS. Patients with Gleason score of 6 ('Gleason 6 patients') were thought of as ideal candidates for AS, although most clinicians set additional limits based on other clinical characteristics. One clinician was distinct from others in his view that all Gleason 6 patients were suitable for AS, regardless of the volume of cancer detected in biopsy:

INT: So if it was a man with Gleason 6, what number of cores would that be coupled with and what volume of core would you be happy with?

U4: Yeah, well most men with Gleason 6 prostate cancer, I would be happy to treat them with active surveillance.

INT: OK, regardless of number of cores?

U4: Yeah.

Other informants consistently reported that the number of cancer cores detected in the biopsy would be an accompanying consideration, irrespective of the low Gleason score:

So we're talking about 1, 2 or perhaps 3 out of the 10 cores that are positive. So if they've got many cores positive, even Gleason 6 we're less likely to do it.

U1

All participating sites included some patients with Gleason score of 7 in their AS programmes. These were generally limited to Gleason 3 + 4 disease. Only one clinician felt that Gleason 4 + 3 disease would be appropriate for AS. Although the clinician appeared to speak on behalf of their clinical institution, none of the other clinicians from this site expressed this view:

I mean obviously 3 + 4 you're taking less of a risk. Um but I think it's reasonable to do a 4 + 3.

U10

All clinicians portrayed Gleason 3 + 4 disease to be in a higher-risk category than Gleason 6, as reflected through their specification that other factors would need to compensate for the higher score:

I always feel a little bit uneasy about it actually but they have to be low volume, definitely low volume and I think they really need to be T1C really.

U3

There's the odd patient with a Gleason 7, but with a tiny amount, maybe.

U7

Four out of the 17 professionals interviewed expressed unease with monitoring Gleason 7 patients (irrespective of other criteria). These views were expressed across three of the four sites, by one nurse and three urologists. As explained by the specialist nurse below, this discomfort was related to the belief that there was greater scope for error and missed opportunities in treating this higher-risk group of patients:

Yes, yes just because the 7s are um – you don't know how they're going to change and what effect delaying more radical treatment may have and, you know, is there a possibility that you would miss the boat?

SPN3

Volume of cancer

The volume of cancer detected through biopsy featured in all clinicians' descriptions of ideal AS candidates. This was often simply phrased as 'low-volume' disease, but clinicians found it difficult to apply numerical limits. The concept of volume appeared to involve balancing the number of positive cores detected against the extent to which each core had cancer present. A higher figure for one of these measures could be compensated by lower values in the other:

Essentially if we were going to tie it down we would say no more than three cores, no core more than 50% Gleason 3 plus 3 and then we broaden a bit by including some 3 plus 4s, occasionally might let somebody in, there was tiny amounts in four cores for example, we might take somebody who had a single core on the out of 10 that had 60% . . .

U1

Clinicians' reluctance to give absolute figures was a reflection of how they operated in practice. Other factors, both clinical and patient-related, could also influence the upper thresholds for volume. The clinician below explained how discomfort about a patient having five positive cores could be alleviated if other factors were favourable:

U5: Once it gets up to five I start feeling very uncomfortable about not having radical treatment- but if they had a low PSA and if the percentage involvement of the cores is low – so you could have someone with four cores, but only 5% of each of the cores, it's still quite a low volume – as long as they've been . . . as long as all the other factors are favourable and they were being monitored closely and the patient is aware that they're in a slightly higher risk category than someone with just one core, then I would go along with it because it's partly up to the patient.

INT: In terms of the percentage per core . . . what's your kind of threshold?

U5: It depends how many cores – it's kind of like the overall picture rather than one specific thing isn't it, so it's difficult to answer that, but you know if it was just one core with a high percentage volume then I'd be happy for active surveillance, if it was four cores with a . . .

INT: So how high . . .

U3: Oh over 50 I would say is high . . . if it's sort of four cores with 50% volume, I'd be a little bit more inclined towards radical treatment than someone who had four cores with 10% volume you know.

Prostate-specific antigen values

Most clinicians had a preference for newly diagnosed patients to have a PSA value < 10 ng/ml. This appeared to be driven by guidelines rather than clinicians' own beliefs about PSA. For example, most of these clinicians talked about PSA levels needing to be < 10 ng/ml for the patient to be considered low risk, based on their recollections of risk stratifications published in official guidance (e.g. NICE,⁶⁸ British Association of Urological Surgeons [BAUS] and European Association of Urology guidance):

INT: You know you mentioned that you would want men with a PSA less than 10?

U15: Yeah.

INT: Is there any flexibility with that? What is it about 10?

U15: Ten? It probably is evidence-based, but 10, I mean that's the criteria, the BAUS criteria for low risk. If it's 10 to 20, that puts you up into intermediate risk. So the NICE guidelines and BAUS recommendations would be that active surveillance is an option, but it's not the first option. Whereas with 'low risk', active surveillance is the first option.

Despite this, the vast majority of clinicians presented the threshold of 10 ng/ml as a preference rather than a set rule. In practice, the patient's PSA value would be one of many factors considered. The most frequently discussed measures among this group was that of prostate volume:

INT: You mentioned PSA of 10: what's the significance of the 10 value?

SPN2: Well again that is pretty subjective, and it depends on which piece of literature you use. I think if you read the guidelines that do exist from various places, they kind of use 10 as being a cut-off, but obviously also it depends on the gland volume. And so, you know, you could have a gentleman who might have a single core, so very small volume disease, but he might have a really big prostate and his PSA might be 10 or 11, so you've got higher PSA level, but that, you know, you could expect that to be a benign element of it, as opposed to necessarily to do with the prostate cancer.

Approximately half of clinicians did not provide a threshold for PSA, citing prostate volume and individual variations as reasons for this. PSA values were seen as one of many components of the bigger picture:

Well, I'm sorry to be, it's very difficult, you see, because if someone's got a large prostate, they will have a larger PSA because of their prostate volume [. . .]. So you can use a PSA over the prostate volume, which we don't do officially, but you do it mentally in your own, yourself in the clinic, when you're with that patient – and it's just a case of, as with all these things, you build up a picture. You feel what the prostate feels like, you look at the patient – how well they are, what other problems they have, what the biopsies show, how big the prostate is, and these things . . . and this is how we make the decisions in real life.

U6

Patient age

Age of patient at diagnosis was another factor that contributed to clinicians' decisions to suggest AS as an option. AS could only be offered if potentially curative treatment was a safe option for patients. As such, patients could not be biologically (rather than chronologically) too old:

It doesn't necessarily mean that age, it does figure into it but if you're 85 you're not going to benefit from surgery. But you know if you're fit, 75/76 year old, have still got a 10-year life expectancy so you know surgery would still be an option.

U15

Making decisions about particularly young patients (i.e. men in their 40s and 50s) was more likely to provoke anxiety in clinicians. Some talked about the need to strike a balance between preserving quality of life (i.e. continence and sexual function) yet not 'missing the window' for treatment:

You've got to look at the co-morbidities and of course their age. So it's less suitable the younger the man is because they've got a longer life expectancy and the outcome is much less certain.

U10

So the younger a person, the more worrying it is, but perhaps the benefits from delaying treatment might be greater as well. [. . .] They may reckon if they can live with the cancer for a couple of years and maintain sexual function that is better for their quality of life, and then that's very reasonable, but of course there is the risk of a disease progressing. [Later] Younger people's diseases behave worse.

U9

One clinician was distinct in feeling that the younger age groups were 'perfect' candidates for AS owing to the quality-of-life implications of treatment side effects. There was no mention of these patients being high risk or unpredictable:

[. . .] the best people for active surveillance are the younger men. Again, contrary you will hear people say they won't consider young men, as I just said, I started measuring my PSA at the age of 50 because I think I'm a perfect candidate for Active Surveillance because I don't want to be impotent unless I have to, so if I get low-volume, low-grade disease I want Active Surveillance at 50 or thereabouts.

U1

Men's lead-up to diagnosis

All men described receiving their diagnosis within a secondary care clinic following the results of a biopsy.

Twelve of the 20 men had been referred to secondary care following disclosure of urological symptoms to their GP. This was the primary reason for consulting for seven men, all of whom were offered a PSA test. Of these, one man described requesting this:

P2: Well, for some time I'd had problems going to the toilet. So I thought, as soon as I, PSA, I decided that once it started to [. . .] be a slight concern to me I had the PSA done.

INT: Right. So it was your choice?

P2: My choice, yeah. I went to the doctor and asked for the PSA tests.

Four of the 11 men had consulted for other reasons, but mentioned their urological symptoms as a secondary issue.

Three men described how various health-care providers had suggested the PSA test. Two of these three men were referred to secondary care and diagnosed with PCa following their first PSA test. One of these recommendations to have the PSA test came from a man's GP, in response to his history of urological infections and his age (over 60 years). Recommendations made to the other two men came from nurses involved in the long-term management of unrelated chronic conditions:

And she said 'Have you had a blood test for your PSA? I said 'No, I've never heard, I don't know what you're talking about. What is it?' She said, 'Well, it's what men get, you know.' She said 'Well we'll give you a blood test.'

P5

Three men had opted for routine PSA testing themselves, on the basis that they were aged over 60 years:

I went to my doctor and just wanted a, really as a check over, you know, a well man clinic. And part of that check-up was a PSA test. It came back slightly raised at 4.8. So they decided to run another PSA test a month later and it had gone up to 5.4. Umm, and the fact that it had gone up again began ringing alarm bells.

P6

One man had started routine PSA testing following recommendations from friends:

He sort of pushed me to go and get it done. So I went for a health check [. . .] I was just over 60, I think, and when I went for a health check I asked for that as well. So the first two times I had it done there was nothing. [. . .] And then I think the third time [. . .] 6.1.

P20

Finally, one man described receiving a PSA test as part of a series of blood tests he had received while receiving hospital care following an accident. Interestingly, this man did not recall being informed of this PSA test at the time:

So, I think it was fortuitous, accidental. I suspect that had I not had the accident [. . .], I probably, even now, would not have known about the prostate.

P1

Clinicians' accounts of active surveillance follow-up protocols

Prostate-specific antigen testing

Clinicians from three sites reported that patients received 3-monthly blood tests via their GPs. Some clinicians suggested that there was flexibility surrounding the 3-monthly PSA testing if dealing with stable patients who had been on AS for a prolonged period (i.e. 18 months to 2 years):

After 18 months or so you can go to 6-monthly.

U15

The consistency of PSA testing could be considered as one of the more fixed components of the AS programme for most sites. There was more variation observed across clinicians' accounts at the site where the protocol was under development, with discrepancies in how regularly PSA was reportedly monitored, and how and when these patterns could change:

Well they um – 3- to 4-monthly PSAs first year, then variable.

SPN3

So following an outpatient discussion on the available evidence that that's what we're going to do, then the patients will have another PSA three months after that. And then depending on the results, there will be a review of the PSA 3 to 6 months after that.

U8

Overall, all clinicians suggested that there would be an initial PSA 3 months into starting an AS programme. Beyond this, reports tended to align with 3–4-monthly PSAs within the first year, followed by a flexible approach:

Um we generally do it for at least a year, I would say probably 2 years, and then we step it down a bit after that. But again, you know, if you've got somebody who is 69 on surveillance you'd probably drop it after a year.

U10

Digital rectal examinations

Clinicians' reported practices of conducting DREs were similar across three of the sites. Nonetheless, clinicians' views on the value of DREs varied by individual, irrespective of which site they were from.

Clinicians from three sites generally reported conducting DREs every 6 months, corresponding to whenever a patient had an outpatient appointment. Two urologists were distinct in conducting annual DREs. The fact that clinicians from one site produced wide-ranging responses was likely to be connected to the absence of a protocol:

INT: How often do you do DREs then?

U8: Ah that varies on whether – I mean if these are – on the basis that usually these are all T1C tumours, so impalpable, um then they will have a DRE when they come back obviously for the first repeat biopsy. And then I will do it afterwards in the outpatients [...] but there isn't a set protocol, the same with the follow-up intervals, they are not prescribed, but I think it is important that one doesn't rely just on the PSA.

In line with the comment above, clinicians across all sites often framed the DREs as a way to compensate for the limitations of PSA measures:

We do occasionally get patients whose PSAs remain stable and we do a rectal, in fact we had one just recently and did a rectal examination and it's a significantly advanced disease and the dynamics have changed dramatically.

U2

Because PSA has been well shown not to be a perfect marker for progression and the DRE sometimes picks things up.

U1

Some clinicians commented that DREs were highly subjective – especially if performed by different clinicians over time:

I think if you're having the same person carrying out the DRE every time, that that's probably more useful. I know that that isn't practical. But you're getting different people's interpretations of, you know, moderately enlarged to one person, you know, what degree is moderately enlarged? And um feeling a nodule had – you know, it's personal interpretation, isn't it? And I think there isn't enough consistency probably.

SPN3

Some clinicians implied that DREs were not sensitive enough to detect change, yet still conducted these regularly for reasons stated earlier (i.e. to avoid relying on PSA alone). In the case of the clinician below, DREs were thought to have psychological benefits for patients:

I think that it's valid psychologically to the patients. I think men find it very reassuring to be examined for something that's not being treated. How much that really is a sensitive test of the clinical progression in somebody with low risk disease in the first year when realistically things tend not to change, I question.

U3

The greatest range of views on DREs came from the site at which the protocol was under development. Quotations demonstrating the disparate views at this site are shown in the main body of the report (see *Chapter 6*). Views ranged from one urologist believing that there was no value in conducting DREs in an AS programme that included rebiopsy, to another urologist implying that examining the prostate was one of the most fundamental procedures that should be carried out when trying to monitor prostate cancer.

Rebiopsy

Rebiopsy provoked the greatest variation of views across clinician interviews, and, for some, the greatest expression of uncertainty. The site protocol was consistently reported by most clinicians from site 1: rebiopsies were offered one year after starting on AS, followed by 2-yearly intervals. As before, clinicians viewed this as an outline, making it clear that there was flexibility within this depending on individual patient cases.

The greatest range of rebiopsy practices emerged from the site at which the protocol was under development. Variation was found in relation to 'initial' rebiopsy times (ranging from 6-monthly to yearly) and subsequent rebiopsies (ranging from yearly to 'based on need'). Clinicians from this site were distinct from others in that they were explicit about the uncertainties over when to rebiopsy (see *Chapter 6, Current practice within active surveillance programmes: follow up procedures and triggers for radical treatment* for quotations that succinctly demonstrate this).

Men's reassurance on active surveillance: clinicians' descriptive terminology of low-risk prostate cancer

At least half of the men interviewed talked about the severity of their cancer by repeating the terms they recalled hearing at the time of diagnosis. Although this might be expected, it was interesting that men offered these recollections of clinicians' descriptions without prompting from the interviewer. This suggested that the information they had gleaned from this consultation played a key part in shaping how they understood their diagnosis. Most made reference to characteristics such as the grade, size, aggressiveness and risk-level of the cancer:

We're bright enough to sort of understand 'low-grade' and 'non-aggressive'.

P1

Now, if I was in deeper water, and things weren't looking quite so rosy, um, maybe I might view it differently. I really don't know. Because I've been told it's a very low-level risk . . .

P3

So I think, crikey: I've been told it's not life-threatening, it's not aggressive.

P7

Some men were able to recall more detailed elaborations of their risk status that had been communicated by their clinicians:

Two samples were showing up . . . But he said ten years ago we wouldn't have taken any notice of that.

P3

Well, the lady registrar said that I'd got it. She said but it's like I've got a – it's like a pussy cat, and hoping it stays like a pussy cat and don't turn into a tiger. That's how she described it.

P5

Although there may have been limitations to portraying accurately what men had been told at the time of diagnosis, two explanations of risk status recurred across numerous interviews. Eight men explained their positive prognosis in terms of the likelihood that they would die with rather than of prostate cancer:

I got the impression that it wasn't necessarily going to be that active or spread that quickly, um but there seemed to be a good chance I'd die of something else before it actually got too serious. That was the sort of conclusion to which I think I arrived. I'm not sure whether that was exactly what they were telling me or just what I wanted to believe, but that seemed to be the message I got.

P14

Another recurring theme was the idea that low-grade PCa is a 'hidden', vastly prevalent condition that goes undetected among the male population. This normalising of low-risk PCa appeared to reassure men. The idea that the man could have been one of the thousands of men unaware of their cancer status may have helped patients to refrain from viewing themselves as cancer patients (or 'ill'):

I think the first time- the first consultant [...] said something like – I don't know – 70% of men who will die of old age have probably got prostate cancer but they don't actually know it [...]. And I just wonder if [...] my doctor hadn't said 'oh let's do a blood test' when I went with my back problem [...], then I guess I wouldn't have known at this moment in time.

P17

Clinicians' early descriptions of men's cancer clearly set the tone for how men went on to understand their condition in this purposeful sample of men with low-risk, low-volume cancer. The power behind clinicians' actions and words was clear in the following explanation of how this patient had come to interpret his diagnosis:

All right, nobody can say, well, don't worry about it, 'cos I know people can't say that when you've just been told you've got cancer. But their body language has always been, well, we found it, but you're a long way, and it's a non-aggressive type [...] I think if they could, they'd say, 'don't worry about it'. But I don't think they can ever really say that to anybody, because who's not going to worry about having cancer in some form or another? But I mean they've all pretty much said that one way or another, ain't they, that [Partner: Mm] we keep an eye on you, but said not to worry about it.

P7

Patients' perceptions of clinicians' triggers for considering radical treatment

Men were asked to describe any changes in PSA that would concern them, but all responses were framed around explanations they recalled hearing from their clinicians. The most common response involved rises in PSA, described as a series of consecutive rises, or a steeply rising gradient of PSA figures:

And they've told me that one bad, one high figure, doesn't mean anything. They wouldn't react to [...] one high PSA figure. It's only a, you know, an increasing pattern that they'll respond to.

P8

Well I got the impression that if it started . . . I think that if it was rising fast, then I should be considering treatment.

P13

Three of 20 men talked about the PSA value of 10 ng/ml being a significant threshold, although each of these qualified this by saying that the trend in PSA values was more important than absolute figures. All three felt that clinicians had indicated there was some significance to a PSA value of 10 ng/ml. One of these men felt that this might have been an outdated idea:

I can remember from years ago that 10 used to be a point where they were started to get a bit panicky about it and think that perhaps they ought to do something about it. It sort of stuck in my mind. Now, I don't think they're quite concerned so much about that now. I think they're more interested in a trend for you as a person, rather than 10 as a nasty point.

P2

Other individual examples of personal triggers for concern included a doubling of the most recent PSA figure (P1), and deviations from age-adjusted reference lines for PSA (P4). The men who mentioned these triggers had based this on discussion with their clinicians. P1 recalled asking his consultant what a 'significant change' in PSA value would be. P4 was the only patient to report being given age-adjusted reference lines that denoted average PSA values in men without PCa. Hovering slightly above this line was deemed acceptable to this man, in light of his PCa diagnosis:

In other words, I'm behaving like a mid-70-year-old in my late sixties. So, that bit of information has been very helpful to me, to put it into perspective.

P4

None of the men reported that he would look into radical treatment in the face of concerning PSA values. All participants responded that the next step would be to initiate a discussion with their consultant (and in one case, their GP), or to wait for the results of a rebiopsy.

Quality assurance checklist for reporting of qualitative research

The reporting of the qualitative methods were checked against the Consolidated criteria for Reporting Qualitative research (COREQ), a 32-item checklist for interviews and focus groups.¹²² The checklist is shown below; the final column details where evidence can be found that the relevant criterion is satisfied.

TABLE 23 Quality assurance checklist for reporting of qualitative research

Number	Item	Guide questions/description	Cross-reference where included in report/reason for non-inclusion
Domain 1: Research team and reflexivity			
<i>Personal characteristics</i>			
1	Interviewer/facilitator	Which author(s) conducted the interview or focus group?	See <i>Data collection methods</i>
2	Credentials	What were the researcher's credentials (e.g. Doctor of Philosophy, Doctor of Medicine)?	See <i>Qualitative researcher profiles</i> , below this table
3	Occupation	What was their occupation at the time of the study?	See <i>Qualitative researcher profiles</i> , below this table
4	Gender	Was the researcher male or female?	See <i>Qualitative researcher profiles</i> , below this table
5	Experience and training	What experience or training did the researcher have?	See <i>Qualitative researcher profiles</i> , below this table
continued			

TABLE 23 Quality assurance checklist for reporting of qualitative research (continued)

Number	Item	Guide questions/description	Cross-reference where included in report/reason for non-inclusion
<i>Relationship with participants</i>			
6	Relationship established	Was a relationship established prior to study commencement?	See <i>Prior relationship/contact with research participants</i> , below this table
7	Participant knowledge of the interviewer	What did the participants know about the researcher (e.g. personal goals, reasons for doing the research)?	See <i>Prior relationship/contact with research participants</i> , below this table
8	Interviewer characteristics	What characteristics were reported about the interviewer/facilitator (e.g. bias, assumptions, reasons and interests in the research topic)?	See <i>Prior relationship/contact with research participants</i> , below this table
Domain 2: study design			
<i>Theoretical framework</i>			
9	Methodological orientation and theory	What methodological orientation was stated to underpin the study (e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis)?	See section <i>Chapter 6, Analysis</i>
<i>Participant selection</i>			
10	Sampling	How were participants selected (e.g. purposive, convenience, consecutive, snowball)?	See <i>Sampling and recruitment of clinicians/nurses</i> and <i>Sampling and recruitment of men on AS</i>
11	Method of approach	How were participants approached (e.g. face-to-face, telephone, mail, e-mail)?	See <i>Sampling and recruitment of clinicians/nurses</i> and <i>Sampling and recruitment of men on AS</i>
12	Sample size	How many participants were in the study?	See <i>Participant characteristics</i>
13	Non-participation	How many people refused to participate or dropped out? Reasons?	See <i>Participant characteristics</i>
<i>Setting</i>			
14	Setting of data collection	Where were the data collected (e.g. home, clinic, workplace)?	See <i>Chapter 6, Data collection methods</i>
15	Presence of non-participants	Was anyone else present besides the participants and researchers?	See <i>Chapter 6, Data collection methods</i>
16	Description of sample	What are the important characteristics of the sample (e.g. demographic data, date)?	See <i>Participant characteristics</i>
<i>Data collection</i>			
17	Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	See <i>Chapter 6, Data collection methods</i>
18	Repeat interviews	Were repeat interviews carried out? If yes, how many?	See <i>Chapter 6, Data collection methods</i>
19	Audio-/visual-recording	Did the research use audio- or visual-recording to collect the data?	See <i>Chapter 6, Data collection methods</i>
20	Field notes	Were field notes made during and/or after the interview or focus group?	See <i>Chapter 6, Data collection methods</i>
21	Duration	What was the duration of the interviews or focus group?	See <i>Chapter 6, Data collection methods</i>

TABLE 23 Quality assurance checklist for reporting of qualitative research (*continued*)

Number	Item	Guide questions/description	Cross-reference where included in report/reason for non-inclusion
22	Data saturation	Was data saturation discussed?	See <i>Chapter 6, Data collection methods</i>
23	Transcripts returned	Were transcripts returned to participants for comment and/or correction?	Transcripts were not shared with participants
Domain 3: analysis and findings			
<i>Data analysis</i>			
24	Number of data coders	How many data coders coded the data?	See <i>Chapter 6, Analysis</i>
25	Description of the coding tree	Did authors provide a description of the coding tree?	See <i>Chapter 6, Analysis</i>
26	Derivation of themes	Were themes identified in advance or derived from the data?	See <i>Chapter 6, Analysis</i>
27	Software	What software, if applicable, was used to manage the data?	See <i>Chapter 6, Analysis</i>
28	Participant checking	Did participants provide feedback on the findings?	A lay summary of findings was shared with participants, although this was not for purposes of 'participant checking'
<i>Reporting</i>			
29	Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified (e.g. participant number)?	Evidenced throughout <i>Chapters 6 and 7</i>
30	Data and findings consistent	Was there consistency between the data presented and the findings?	Evidenced throughout <i>Chapters 6 and 7</i>
31	Clarity of major themes	Were major themes clearly presented in the findings?	Evidenced throughout <i>Chapters 6 and 7</i>
32	Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Evidenced throughout <i>Chapters 6 and 7</i>

Qualitative researcher profiles

LR and JW are both trained (female) qualitative researchers who hold doctorates in health services research. At the time of interviews, LR had one year of postdoctoral experience, whereas JW had 5 whole-time-equivalent years of postdoctoral experience. Prior to their research careers, LR was an undergraduate student studying physiological sciences and JW worked as a qualified speech and language therapist.

LR approached this project as a naive researcher, with little knowledge of the field or literature. JW had prior experience of working on the ProtecT trial as a qualitative researcher and thus had extensive knowledge of the field through her awareness of the literature and her own research with men diagnosed with PCa (in the context of the ProtecT trial). This prior knowledge may have influenced JW's interpretation of data, although JW's contributions were largely based on LR's initial interpretation and analytical thoughts, rather than raw data. We have no reason to believe that either researcher's personal experiences or life history would have influenced their approaches to data collection and analysis in a noteworthy way.

Prior relationship/contact with research participants

Neither LR nor JW had prior contact or relationships with the research participants, bar clinicians whom they had met at study introductory meetings (where a number of clinicians were recruited).

Clinicians and men all received study information sheets in advance of the research interview, which explained why the research was being undertaken. Information from these sheets was condensed into a brief slide presentation (Microsoft PowerPoint®; Microsoft Corporation, Redmond, WA, USA), which was shown to some clinicians at study introductory meetings.

Information provided at the outset, either via information sheets or presentations, included no personal details about the researchers or their personal aspirations. As such, interview participants will not have been informed of any personal information about the researchers prior to being interviewed. A number of men asked LR about her career and history after interviews (in the context of informal conversation).

Appendix 4 Description of statistical methods used to model repeated measures of prostate-specific antigen

Linear mixed models

A LMM^{37,123} for the observed responses Y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, n_i$, where i indexes individuals and j indexes measurements within individuals, can be written as:

$$Y_{ij} = W_i\beta + X_{ij}\gamma + Z_iu_i + \varepsilon_{ij}, \quad (9)$$

where $\beta = \beta_0, \dots, \beta_p$ is the $p + 1$ dimensional vector of fixed effects and $\gamma = (\gamma_1, \dots, \gamma_q)$ is the q dimensional vector of time effects. The random effects $u_i \sim N(0, \Sigma_u)$ are assumed here to have an unstructured $r \times r$ covariance matrix Σ_u . W_i , X_i and Z_i are the fixed, time and random design matrices, respectively, which represent the fixed, time and random covariates specified in the model. The $\varepsilon_{ij} \sim N(0, \Sigma_\varepsilon)$ are the residuals from the model. This model assumes that the relationship between outcome and covariate is linear. This may not always be the case, and so we discuss two methods for accommodating non-linear relationships within the mixed-models framework.

Fractional polynomials

Royston and Altman¹²⁴ suggest a simple set $S = (-2, -1, -0.5, 0, 0.5, 1, 2, 3)$ of powers for transformation of a covariate in a multiple regression setting (the power 0 is taken as $\log [X]$). Fractional polynomials (FPs) are used to model a response which is non-linear in some covariates. FP1 denotes a first degree fractional polynomial where a transformation of a covariate X is captured in a single term (i.e. X^q , $q \in S$; for a single covariate X). A second degree fractional polynomial (FP2) transforms a covariate X as $X^q = X^{(q_1, q_2)}$ where:

$$X^q = X^{(q_1, q_2)} = \begin{cases} (X^{q_1}, X^{q_2}); & q_1 \neq q_2 \\ (X^{q_1}, X^{q_1} \log X); & q_1 = q_2 \end{cases}. \quad (10)$$

The best FP degree and power(s) are found through first fitting all models for each degree and choosing the model from each degree (i.e. FP1, FP2) with the lowest deviance. Once the 'best' model of each degree has been selected, a closed form algorithm¹²⁵ is used to find the best model. This algorithm provides a straightforward hypothesis test of the linearity assumption, which should be checked before continuing with a linear model.

Fractional polynomials were developed primarily for standard regression analysis but are easily commuted to a multilevel framework.^{98,126} The model is written as in (1) where the time coefficient $\gamma = (\gamma_1, \dots, \gamma_m)$ is a vector of length m , corresponding to the degree of fractional polynomial used. For example, a FP2 mixed model, taking powers 1 and 3 from S , with no fixed covariates W , would be:

$$Y_{ij} = \beta_0 + X_{ij}\gamma_1 + X_{ij}^3\gamma_2 + u_{0i} + X_{ij}u_{1i} + X_{ij}^3u_{2i} + \varepsilon_{ij}. \quad (11)$$

By incorporating a random intercept and random effects for all terms involving time (X_{ij}), the individuals may vary about the typical intercept and polynomial trend (linear and cubic trend here).

Regression splines

Splines are piecewise polynomial functions that allow the response to be modelled differently in separate intervals of the time covariate X . This is done by introducing knots or breakpoints to partition the range of X . In statistical analyses, spline functions may offer the flexibility required to describe accurately non-linear patterns which may exist in the relationship between variables. A linear spline basis for a single covariate leads to the so-called broken stick model, where the fit is $K + 1$ line segments joined end to end at the K knots. Only the linear spline model is considered here, but higher degree spline bases may be used to capture a conceptually smoother process.

Regression spline mixed models are simply LMMs with a reparameterisation of the time covariates.^{127–129} The model can be written as (1) with time coefficients $\gamma = (\gamma_1, \dots, \gamma_S)$, where S is the sum of the degree of spline basis and the number of knots used. For example, taking one knot at time τ for all individuals, a RSMM with no fixed covariates W is:

$$Y_{ij} = \beta_0 + X_{ij}\gamma_1 + (X_{ij}-\tau)_+\gamma_2 + u_{0i} + X_{ij}u_{1i} + (X_{ij}-\tau)_+u_{2i} + \varepsilon_{ij}, \quad (12)$$

where $a_+ = a$ if $a > 0$ and 0 otherwise. This model allows each individual to have their own intercept ($\beta_0 + u_{0i}$), their own slope ($\gamma_1 + u_{1i}$) and their own adjustment to this slope after the knot ($\gamma_2 + u_{2i}$). Regression splines also offer a framework for checking the assumption of linearity of the response in the predictor. If any estimated spline coefficients are found to have CIs not containing 0, this gives evidence against the linearity assumption.

Functional principal components analysis

Functional data analysis¹³⁰ offers an extension of non-parametric smoothing to repeated measures data. These methods allow for flexible curves to be fitted to each member of a group and are very useful when forgoing any assumptions about the shape of these curves. However, for a large number of individuals with irregular measurement times, several functional data analysis methods become inefficient. Yao *et al.*¹³¹ propose a version of functional principal components analysis (FPCA) whereby sparse and irregular longitudinal data (such as those in our example) can be modelled.

The process of FPCA is highlighted in *Figure 14* (constructed using the PACE example in MATLAB). Firstly, the hierarchical structure of the data is ignored and a smooth curve is fitted to the pooled data. This estimate for the mean is then used to build up a matrix of covariances; these represent the deviations from the mean for each pair of time points. For example, two residuals on the same side of the mean fitted curve would have a positive covariance; two on opposite sides would have a negative covariance. In order to construct the required two-dimensional curves, this three-dimensional covariance matrix (or surface) is decomposed (or summarised) into a linear combination of orthogonal eigenfunctions and eigenvalues. These eigenfunctions, known as functional principal components (FPCs), act as a basis on which individual trajectories can be constructed. In terms of mixed models, the FPCs can be seen as patterns of within subject variance left over after the mean fit. The first FPC summarises the main pattern of variation from the mean. After this, the second FPC, which is orthogonal to the first, explains the next main pattern of variation from the mean and so on. Thus, the data are now summarised in terms of the mean pattern and functions of the variation from the mean. The final step is to construct individual curves using these summaries of the complex data. The individuals' curves are found by multiplying the FPCs by scaling factors which quantify the extent to which the individual's trajectory correlates with the corresponding FPC (or pattern of variation). For instance, if an individual follows the mean pattern exactly, their FPC scores would be zero. The combination of the mean fit and FPC score scaled individually results in a smoothed curve for each member of the cohort.

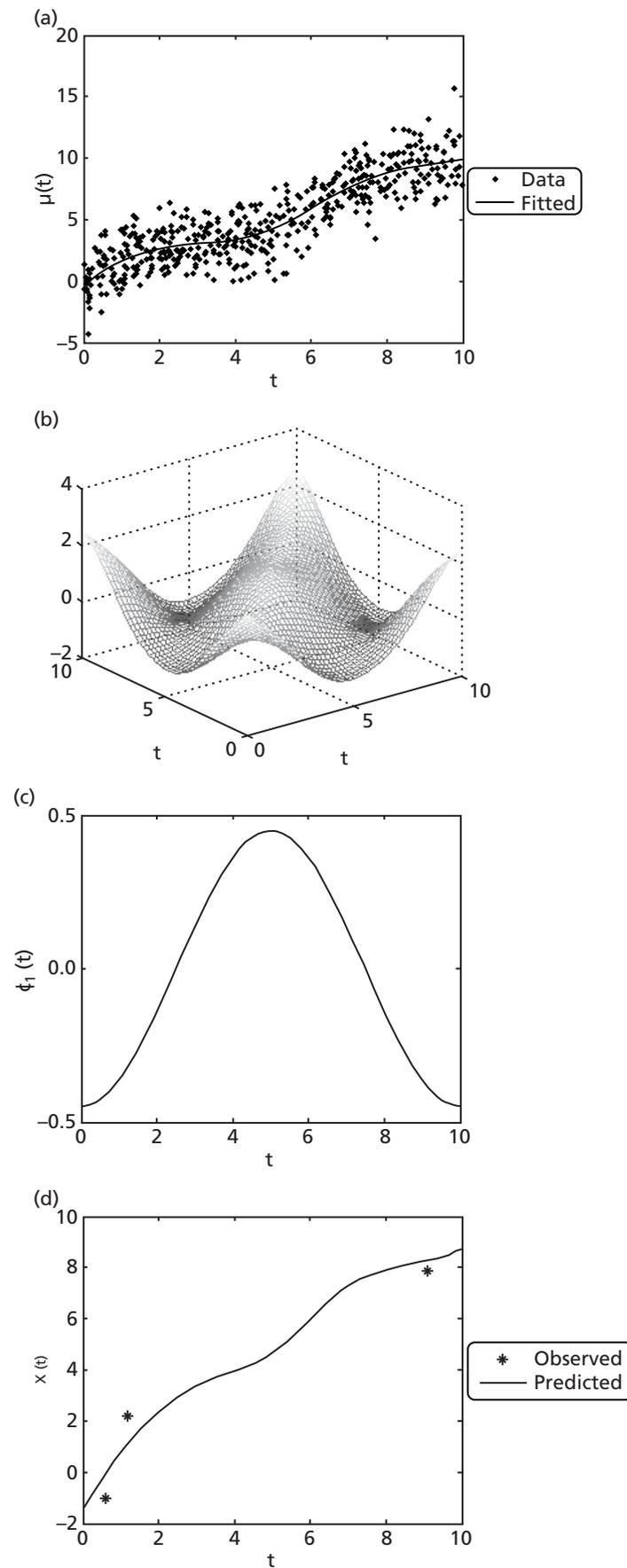


FIGURE 14 Fitting process of functional PCA. (a) Smooth all data; (b) smooth covariance surface; (c) find FPCs from smoothed surface; and (d) obtain individual curves using mean fit and FPCs.

Appendix 5 Description of the studies and data used for analysis

TABLE 24 Description of the studies used for analysis

Cohort	ProtecT	RMH	JH	UCHC	SPCG4
Number of men	512	499	961	114	290
Years of diagnosis	1999–2009	1999–2010	1992–2012	1990–3	1989–99
Years of PSA testing	1999–2011	1999–2012	1992–2012	1990–2005	1989–2005
Inclusion criteria	Clinically localised PCa were offered randomisation	Baseline PSA < 15 ng/ml; Gleason score of ≤ 3 + 4; T2; percentage of positive biopsy cores ≤ 50%	PSAD < 0.15 ng/ml/cm ³ ; Gleason score of ≤ 3 + 3; T1c; two or fewer positive biopsy cores, with maximum involvement of 50% per core	NR	Age < 75 years; T0d, T1 or T2; life expectancy > 10 years
Monitoring schedule	PSA tests every 6 months	PSA tests every 3–4 months in the first 2 years then every 6 months	PSA tests every 6 months	NR	NR
PSA testing in underlying population	Common: but in this trial all were screened	Common	Widespread	Rare	Rare
Clinical exposures available	Gleason score, T-stage, percentage of positive cores, prostate volume, percentage-free PSA	Gleason score, T-stage, percentage of positive cores, prostate volume, percentage-free PSA	Gleason score, percentage of positive cores, percentage of cancer per core, prostate volume, PSAD, percentage-free PSA	Gleason score	Gleason score, T-stage
Clinical outcomes available	NR	Metastases, PCSM	PCSM, all-cause mortality	All-cause mortality	Metastases, PCSM

NR, not recorded.

TABLE 25 Characteristics of the data received

Cohort	ProtecT	RMH	JH	UCHC	SPCG4
Number of men	512	499	961	114	290
Number of PSA tests	7438	9427	9993	884	2987
Average age at diagnosis, years (SD)	NR	66.0 (6.2)	65.8 (6.1)	69.8 (4.5)	64.6 (4.9)
Average PSA at diagnosis, ng/ml (SD)	NR	6.87 (3.5)	3.81 (2.8)	9.16 (8.4)	15.0 (11.5)
Average duration of monitoring (SD)	4.8 (2.4)	4.5 (2.6)	3.2 (2.7)	4.7 (3.9)	6.0 (3.8)
Gleason score of 3 + 3, %	NR	91	99.8	75	67
Clinical exposures available	Gleason score, T-stage, percentage of positive cores, prostate volume	Gleason score, T-stage, percentage of positive cores, prostate volume, percentage-free PSA	Gleason score, percentage of positive cores, percentage of cancer per core, prostate volume, PSAD, percentage-free PSA	Gleason score	Gleason score, T-stage
Clinical outcomes available	NR	Metastases, PCSM	PCSM, all-cause mortality	All-cause mortality	Metastases, PCSM, all-cause mortality

NR, not recorded.

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME
HS&DR
HTA
PGfAR
PHR**

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library