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 Neal[6] and Kate Tilling[1]*

[1]School of Social and Community Medicine, University of Bristol, Bristol, UK
[2]Division of Urology, University of Connecticut Health Center, Farmington, CT, USA
[3]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
[4]Faculty of Life Sciences & Medicine, King's College London, London, UK
[5]Regional Cancer Centre, Uppsala/Örebro Region, Uppsala, Sweden
[6]Department of Oncology, University of Cambridge, Cambridge, UK

*Corresponding author

Declared competing interests of authors: none

Published July 2015
DOI: 10.3310/hsdr03300

Scientific summary

Active monitoring of men with clinically localised prostate cancer

Health Services and Delivery Research 2015; Vol. 3: No. 30
DOI: 10.3310/hsdr03300

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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 PSAv > 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/cm^3, ≤ 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 PSAv 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 PSAv > 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 PSAv 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 PSAv. 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.

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.

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