

Keywords: active surveillance; prostate cancer; PSA screening; risk stratification

Identification of pathologically insignificant prostate cancer is not accurate in unscreened men

G L Shaw^{*,1,2}, B C Thomas², S N Dawson¹, G Srivastava², S L Vowler¹, V J Gnanapragasam², N C Shah², A Y Warren² and D E Neal^{1,2}

¹Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK and ²Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ, UK

Background: Identification of men harbouring insignificant prostate cancer (PC) is important in selecting patients for active surveillance. Tools have been developed in PSA-screened populations to identify such men based on clinical and biopsy parameters.

Methods: Prospectively collected case series of 848 patients was treated with radical prostatectomy between July 2007 and October 2011 at an English tertiary care centre. Tumour volume was assessed by pathological examination. For each tool, receiver operator characteristics were calculated for predicting insignificant disease by three different criteria and the area under each curve compared. Comparison of accuracy in screened and unscreened populations was performed.

Results: Of 848 patients, 415 had Gleason 3 + 3 disease on biopsy. Of these, 32.0% had extra-prostatic extension and 50.2% were upgraded. One had positive lymph nodes. Two hundred and six (24% of cohort) were D'Amico low risk. Of these, 143 had more than two biopsy cores involved. None of the tools evaluated has adequate discriminative power in predicting insignificant tumour burden. Accuracy is low in PSA-screened and -unscreened populations.

Conclusions: In our unscreened population, tools designed to identify insignificant PC are inaccurate. Detection of a wider size range of prostate tumours in the unscreened may contribute to relative inaccuracy.

Clinically insignificant prostate cancer (PC) can be defined as a cancer, which will not affect the patient during the natural course of his lifetime. The indolent course of localised low-grade PC means that active surveillance (AS) is considered a good treatment option (Albertsen *et al*, 2005). Recent data regarding the apparent overtreatment of screening detected PC has provoked much debate regarding the way low-risk PC is treated. Generalisation of conclusions from studies where PSA screening is common to populations where PSA testing is relatively uncommon is misleading.

Attempts have been made to define clinically insignificant PC, based on pathological examination of radical prostatectomy specimens and analysis of recurrence rates of these tumours. Several definitions have been formulated. Work by Stamey *et al* (1993) identified organ-confined tumours of $<0.5\text{ cm}^3$ Gleason 3 + 3 with

no grade 4 or 5 as insignificant. Subsequently, Wolters *et al* (2011) analysed the data from the European Randomised Study of screening for PC (ERSPC) (van den Bergh *et al*, 2009) to redefine insignificant PC as organ-confined Gleason 3 + 3 tumours, with no grade 4 or 5, the largest tumour having a volume $\leq 1.3\text{ cm}^3$ and a total tumour volume of $\leq 2.5\text{ cm}^3$. Evidence of tumour volume as an independent predictor of outcome in PC is conflicting, and therefore we considered definitions of insignificant PC that include and exclude tumour volume.

A number of tools have been developed to identify men who are harbouring insignificant PC. The tools are based on parameters defined by clinical examination, serum PSA, transrectal ultrasound findings and biopsy results. These tools have been generated in populations where PSA screening is common from USA (Soloway

*Correspondence: GL Shaw; E-mail: greg.shaw@cruk.cam.ac.uk

Received 28 January 2014; revised 11 March 2014; accepted 15 March 2014; published online 10 April 2014

© 2014 Cancer Research UK. All rights reserved 0007 – 0920/14

et al, 2010; Adamy *et al*, 2011; Tosoian *et al*, 2011; Whitson *et al*, 2011) and from the ERSPC study (van den Bergh *et al*, 2009). In the UK, the uptake of PSA testing is ~6% (Melia *et al*, 2004; Parker *et al*, 2006). We sought to evaluate the accuracy of these tools in identifying insignificant PC (by three different definitions) in a cohort of patients undergoing radical prostatectomy.

Objectives. To evaluate the accuracy of tools designed to identify insignificant PC in a cohort of patients undergoing radical prostatectomy in an unscreened population.

Second, we sought to evaluate the effect of PSA screening on the accuracy of these tools.

MATERIALS AND METHODS

Data were collected prospectively for 847 patients undergoing robotic radical prostatectomy between July 2007 and October 2011 at Addenbrookes Hospital, Cambridge, UK. The PSA level was measured and clinical stage assigned by the attending urologist according to the 2002 TNM staging system. All patients had their biopsy and radical prostatectomy specimens evaluated at our institution by genitourinary pathologists according to protocols from the Royal College of Pathologists (2009).

Data were reviewed retrospectively and 445 patients who had Gleason 3 + 3 on biopsy were identified. Of these, in 415 patients, a specialist uropathologist had estimated tumour volume in the radical prostatectomy specimen at the time of initial histopathological evaluation using a visual inspection method of assessment of marked areas of tumour on whole-mount haematoxylin and eosin stained sections from the entirely embedded prostates (van der Kwast *et al*, 2011).

The histology sections were divided in halves or quadrants to fit cassettes for paraffin embedding. Pathological tumour stage and Gleason grade were assessed. Tumour areas were marked on each slide and measured by visual estimation, giving a percentage of the total volume of prostate occupied by tumour (van der Kwast *et al*, 2011). Subsequently, the volumes of tumour were calculated by multiplying the percentage occupied area by the gland volume (itself calculated by measurement of diameters of resected prostate specimen $H \times W \times L \times \pi/6$ (MacMahon *et al*, 2009)). For 537 patients, the weight of the prostate specimen with seminal vesicles removed was available. Where diagnostic TRUS was performed at the referring centre, we were not able to calculate PSA density using TRUS volume. TRUS-estimated volume was available for 256 patients. In these patients, we used the calculated volume of the resected prostate to calculate

approximate PSA density. Strong agreement was seen between prostate weight and calculated prostate volume, and between TRUS-estimated volume and calculated prostate volume (data not shown).

Postoperative assessments included physical examination and PSA at 6 weeks, 3, 9 and 12 months and every 6 months thereafter. Biochemical recurrence was defined as PSA > 0.2 ng ml⁻¹.

Five criteria designed to identify insignificant disease were identified (Table 1) and performance compared with D'Amico low-risk criteria (PSA ≤ 10, Gleason 3 + 3, cT1-2a) (D'Amico *et al*, 1998). In addition, the discriminative power of the selection criteria used for a UK AS cohort (Selvadurai *et al*, 2013) and the effect of the number of biopsy cores (Ochiai *et al*, 2005) containing cancer were evaluated.

Three different criteria were used to define insignificant PC:

1. Classical definition = organ-confined tumours of < 0.5 cm³ Gleason 3 + 3 with no Gleason 4 or 5 (Stamey *et al*, 1993).
2. ERSPC definition = organ-confined Gleason 3 + 3 tumours, with no Gleason grade 4 or 5, index tumour volume ≤ 1.3 cm³ and a total tumour volume of ≤ 2.5 cm³ (Wolters *et al*, 2011).
3. Inclusive definition = organ-confined Gleason 3 + 3 tumours, with no Gleason grade 4 or 5.

Receiver operator characteristics (ROC) were plotted for each of the tools in predicting insignificant disease by each of the three definitions and the area under the ROC curve (AUC) calculated. Comparisons with the D'Amico tool were made using the DeLong method (DeLong *et al*, 1988).

The D'Amico risk stratification tool was not designed to facilitate selection of patients for AS but as a means to predict outcome following radical therapy (D'Amico *et al*, 1998). It is included here as a benchmark for comparison with tools designed for this purpose. Its use as a predictor of outcome in other treatment modalities has been taken up in the UK (Graham *et al*, 2008).

Sensitivity, specificity, positive predictive value and negative predictive value were calculated to assess the performance of each tool under each of the three criteria. We also considered two additional tools: 'any criteria', which saw the patient recorded as eligible for AS if any one of the tools indicated he was low risk, and 'all criteria', which saw the patient recorded as eligible for AS only if all of the tools indicated he was low risk.

The tools for identifying patients suitable for AS were examined using univariable Cox regression. Unless otherwise stated, *P*-values were calculated based on two-sided hypothesis tests and are corrected for multiple comparisons using a Bonferroni correction.

Table 1. Components of tools by which to identify patients bearing insignificant PC

Reference	Prostate biopsy result			
	PSA (ng ml ⁻¹)	Number of positive cores	Core involvement	Clinical stage
Tosoian <i>et al</i> , 2011	Density ≤ 0.15	≤ 2	≤ 50	T1
Adamy <i>et al</i> , 2011	≤ 10	≤ 3	≤ 50	T1-2a
van den Bergh <i>et al</i> , 2009	≤ 10 and density < 0.2 ml ⁻¹	≤ 2	—	T1c-2
Whitson <i>et al</i> , 2011	≤ 10	≤ 33% (of at least 6)	≤ 50	T1c-2
Soloway <i>et al</i> , 2010	≤ 15	≤ 2	≤ 50	T1c-2
D'Amico <i>et al</i> , 1998	≤ 10	—	—	T1
Selvadurai <i>et al</i> , 2013	≤ 15	—	≤ 50	T1-2

Abbreviation: PC = prostate cancer.

A cutoff of 0.05 was used to determine statistical significance. Statistical analysis was performed using Stata 11 (StataCorp LP 2009, College Station, TX, USA).

RESULTS

Of the 847 patients treated with robotic radical prostatectomy, 415 had Gleason 3 + 3 disease on their diagnostic biopsy and tumour volume estimated after prostatectomy. Of these, 206 patients had D’Amico low risk disease PC (D’Amico *et al*, 1998). It has been shown that the number of cores on diagnostic biopsy was affected by cancer correlates with tumour volume (Ochiai *et al*, 2005). Only 63 patients undergoing surgery had two or fewer cores containing cancer on diagnostic biopsy (all biopsies had 10–12 cores taken) and these men had expressed a strong desire not to pursue AS after careful counselling about the relative merits of AS.

Two hundred and nine out of 415 (50.3%) had Gleason 4 or 5 PC, 131 out of 415 (31.6%) demonstrated extra-capsular extension and 1 out of 415 (0.2%) had positive nodes at surgery, a rate of misclassification in keeping with other non-USA series (Kim *et al*, 2013).

The selection criteria evaluated vary in their stringency. The more stringent criteria render fewer patients eligible for AS, although the number of patients offered AS for significant disease is decreased. Table 2 demonstrates this principle with Toisiain criteria selecting only 3 of 44 (7%) patients with pT3 tumours for AS and the Whitson criteria 39 of 169 (23%) of eligible patients who had pT3 disease. The less-stringent criteria tended to select patients bearing larger tumours. The receiver operator characteristics are shown in Tables 3 and 4. No significant difference in discriminative power is demonstrated for any tool over D’Amico low-risk criteria other than a significant inferiority for the Selvadurai criteria in identifying insignificant disease by the ERSPC definition.

From a pragmatic point of view, the urologist needs to know how many people identified as having insignificant PC by each criteria will turn out to be harbouring significant disease. Table 5 demonstrates the stringency and discriminative power of each tool.

Iremashvili *et al* (2012) compared accuracy of the five tools in predicting insignificant PC in a cohort of North American patients. Table 6 demonstrates the relative accuracy of these tools in our and the Iremashvili cohorts. The AUC did not exceed 0.7 for any of the

tools evaluated in either the UK or USA cohorts. Consensus suggests that a test with an AUC of <0.7 is of limited discriminative power (DeLong *et al*, 1988). In the USA data, the AUCs approached 0.7 for some of the tools predicting insignificant PC by each of the definitions. In the UK cohort none of the AUCs exceeded 0.6.

One limitation of this study is that this cohort of surgical patients may not necessarily be representative of all patients with Gleason 3 + 3 PC on biopsy. The small number of classically defined insignificant PC (*n* = 18) seen limits our ability to analyse this subgroup.

DISCUSSION

The recent publication of the PIVOT trial data has provoked much debate (Wilt *et al*, 2012). PSA screening is commonplace in the USA (for example, in the PLCO randomised study of PSA screening, at enrolment 44% of the men in the ‘unscreened’ arm and this increased to 77% after 5 years on the study (Andriole *et al*, 2009)). In the UK, PSA testing is performed on an *ad hoc* basis and it is estimated that only 6% of men will have had a PSA test (Melia *et al*, 2004; Parker *et al*, 2006). In the USA, the PIVOT study failed to demonstrate a difference in overall survival for men with low-risk PC treated with radical surgery or observation. Conversely, the SPCG-4 study, from an unscreened Scandinavian population, demonstrated a significant survival benefit to radical prostatectomy (Bill-Axelsson *et al*, 2011). Analyses of PIVOT trial data compared with SPCG-4 data suggest that PSA screening results in over diagnosis and lead-time effect (Xia *et al*, 2013). It has been estimated using the ERSPC data that this lead time is 11 years (Draisma *et al*, 2003). Assuming that tumours grow with time, it is logical that screen-detected PCs will be of lower volume than their non-screen-detected counterparts as demonstrated by our data. Examination of ERSPC data reveals a median tumour size of 0.47 cm³ (Draisma *et al*, 2003). Only 4% of our cohort had tumour size ≤ 0.47 cm³ with a median tumour volume of 3.47 cm³ (see Figure 1).

The accuracy of a diagnostic test can be evaluated by calculating the AUC. The AUC for a random act, like a coin toss, is 0.5. The accepted rule of thumb is that an AUC of <0.7 indicates poor discriminative power; between 0.7 and 0.8 indicates acceptable

Table 2. Pathological tumour characteristics in patients identified as bearing insignificant PC by the selection tools described

	Tosoian	Adamy	Van den Bergh	Whitson	Soloway	Selvadurai	D’Amico	Any criteria	All criteria
Patients, <i>n</i> (%)	44 (10.7)	146 (35.6)	63 (15.4)	169 (41.2)	106 (25.9)	306 (74.6)	206 (49.6)	339 (82.7)	43 (10.5)
Pathological Gleason score, <i>n</i> (%)									
2–6	28 (6.8)	85 (20.7)	41 (10.0)	93 (22.7)	58 (14.1)	151 (36.8)	109 (28.6)	169 (41.2)	27 (6.6)
7	14 (3.4)	57 (13.9)	20 (4.9)	71 (17.3)	44 (10.7)	145 (35.4)	91 (21.8)	160 (39.0)	14 (3.4)
8–10	2 (0.5)	4 (1.0)	2 (0.5)	5 (1.2)	4 (1.0)	10 (2.4)	6 (1.4)	10 (2.4)	2 (0.5)
Pathological stage, <i>n</i> (%)									
≥pT3	3 (0.7)	28 (6.8)	6 (1.5)	39 (9.5)	14 (3.5)	90 (22.0)	71 (17.4)	106 (25.9)	3 (0.7)
≥pT3b	0 (0.0)	2 (0.5)	1 (0.2)	2 (0.5)	2 (0.5)	7 (1.7)	3 (0.7)	7 (1.7)	0 (0.0)
Positive margins, <i>n</i> (%)	3 (0.7)	25(6.1)	7 (1.7)	33 (8.1)	16 (3.9)	58 (14.2)	46 (11.3)	65 (15.9)	3 (0.7)
Tumour volume, cc, median (IQR)	1.7 (1.0–3.1)	2.7 (1.3–4.7)	2.5 (1.1–3.9)	2.9 (1.5–5.3)	2.4 (1.1–4.6)	3.4 (1.8–6.3)	3.0 (1.7–5.4)	3.6 (1.9–6.5)	1.6 (1.0–3.2)
Largest tumour volume (where multiple tumours), cc, median (IQR)	1.3 (0.9–2.3)	2.2 (1.0–3.5)	1.8 (0.8–3.2)	2.5 (1.1–3.8)	1.9 (1.0–3.7)	2.8 (1.5–4.4)	2.7 (1.5–4.4)	3.0 (1.6–5.2)	1.4 (0.7–2.6)
Insignificant cancers, classical definition, <i>n</i> (%)	3 (0.7)	9 (2.2)	3 (0.7)	9 (2.2)	8 (2.0)	12 (2.9)	9 (2.2)	12 (2.9)	3 (0.7)
Insignificant cancers, ERSPC definition, <i>n</i> (%)	17 (4.2)	45 (11.0)	22 (5.4)	47 (11.5)	36 (8.8)	66 (16.1)	51 (12.3)	73 (17.8)	16 (3.9)
Organ-confined low-grade cancer, <i>n</i> (%)	28 (6.8)	81 (19.8)	39 (9.5)	88 (21.5)	57 (13.9)	134 (32.7)	97 (23.4)	147 (35.9)	27 (6.6)

Abbreviations: ERSPC = European Randomised Study of screening for prostate cancer; IQR = interquartile range; PC = prostate cancer.

Table 3. Comparison of the diagnostic accuracy of selection tools in identifying patients bearing insignificant PC defined in three different ways

Approach	Insignificant cancer, classical definition				Insignificant cancer, ERSPC definition				Organ-confined low-grade cancer			
	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
Tosoian	0.17	0.9	0.07	0.96	0.18	0.91	0.39	0.78	0.16	0.93	0.64	0.59
Adamy	0.5	0.65	0.06	0.97	0.46	0.68	0.31	0.8	0.45	0.72	0.56	0.63
Van den Bergh	0.17	0.85	0.05	0.96	0.23	0.87	0.35	0.78	0.22	0.9	0.62	0.6
Whitson	0.5	0.59	0.05	0.96	0.49	0.61	0.28	0.79	0.49	0.65	0.52	0.62
Soloway	0.44	0.75	0.08	0.97	0.37	0.78	0.34	0.8	0.32	0.79	0.54	0.6
Selvadurai	0.67	0.25	0.04	0.94	0.68	0.23	0.22	0.7	0.75	0.26	0.44	0.57
D'Amico	0.61	0.36	0.04	0.95	0.67	0.37	0.25	0.79	0.69	0.4	0.47	0.62
Any criteria	0.67	0.17	0.04	0.92	0.75	0.15	0.22	0.66	0.82	0.17	0.43	0.55
All criteria	0.17	0.9	0.07	0.96	0.17	0.91	0.37	0.78	0.15	0.93	0.63	0.59
Cancer cores ≤2	0.5	0.63	0.06	0.97	0.46	0.65	0.29	0.8	0.44	0.67	0.51	0.61
Very low risk ^a	0.44	0.72	0.07	0.97	0.42	0.75	0.35	0.81	0.38	0.78	0.58	0.62

Abbreviations: NPV = negative predictive value; PC = prostate cancer; PPV = positive predictive value.

^aVery low risk = D'Amico low risk with two or fewer biopsy cores containing cancer.

Table 4. Comparison of the areas under the receiver operator characteristic curves of the selection tools

Approach	Insignificant cancer, classical definition			Insignificant cancer, ERSPC definition			Organ-confined low-grade cancer		
	AUC	95% confidence interval	Comparison with D'Amico approach ^a , adjusted P-value	AUC	95% confidence interval	Comparison with D'Amico approach ^a , adjusted P-value	AUC	95% confidence interval	Comparison with D'Amico approach ^a , adjusted P-value
Tosoian	0.53	0.44–0.62	1	0.54	0.50–0.59	1	0.54	0.51–0.58	1
Adamy	0.58	0.45–0.70	0.26	0.57	0.51–0.63	0.462	0.59	0.54–0.63	0.652
Van den Bergh	0.51	0.42–0.60	1	0.55	0.50–0.59	1	0.56	0.52–0.59	1
Whitson	0.55	0.43–0.67	1	0.55	0.49–0.60	1	0.57	0.52–0.62	1
Soloway	0.6	0.48–0.72	0.76	0.57	0.52–0.63	0.882	0.55	0.51–0.60	1
Selvadurai	0.46	0.34–0.57	1	0.46	0.41–0.51	0.024	0.5	0.46–0.55	0.418
D'Amico	0.49	0.37–0.61	—	0.52	0.47–0.58	—	0.55	0.50–0.59	—
Any criteria	0.42	0.30–0.53	0.171	0.45	0.40–0.50	0.001	0.5	0.46–0.53	0.084
All criteria	0.53	0.44–0.62	1	0.54	0.50–0.58	1	0.54	0.51–0.57	1
Cancer cores ≤2	0.56	0.44–0.69	1	0.56	0.50–0.61	1	0.56	0.51–0.60	1
Very low risk	0.58	0.46–0.70	0.418	0.58	0.53–0.64	0.105	0.58	0.54–0.63	1

Abbreviations: AUC = area under the receiver operator characteristic curve; ERSPC = European Randomised Study of screening for prostate cancer. P-values <0.05 are shown in bold.

^aNull hypothesis: AUC_i = AUC D'Amico, two-sided alternative hypothesis: AUC_i ≠ AUC D'Amico. Comparison made using the DeLong method (DeLong *et al*, 1988).

discrimination; and >0.8 indicates excellent discriminative ability (DeLong *et al*, 1988).

In 2012, Iremashvili *et al* (2012) published data examining the accuracy of protocols for selecting patients for AS in a USA cohort. The AUCs ranged from 0.61 to 0.68 for classically defined insignificant PC, 0.59 to 0.64 for updated definition and 0.55 to 0.62 for organ-confined low-grade disease. This compares with 0.51 to 0.60, 0.54 to 0.57 and 0.54 to 0.59, respectively, in our cohort. None of the selection tools evaluated had an AUC of >0.7 in either cohort. There was a trend towards the AUCs being lower in the UK cohort compared with corresponding USA data. We hypothesise that tools based on number or length of biopsy cores

are rendered inaccurate due to greater variability in the size of PCs detected.

Two or fewer cancer-containing cores has been shown to correlate with tumour volume in US series (Ochiai *et al*, 2005). Selecting only patients with two or fewer cores on biopsy or very low-risk disease (D'Amico low risk with two or fewer cores effected) did not improve accuracy in detecting insignificant disease over D'Amico risk stratification alone (Tables 3 and 4).

The recent recommendations of the US preventative task force on PC screening (Moyer, 2012) and the AUA (Carter *et al*, 2013) may result in changes in the uptake of PSA screening worldwide with potential effects on the accuracy of these tools.

Table 5. Number of patients identified correctly and incorrectly as bearing insignificant PC and therefore being eligible for active surveillance

	Tosoian	Adamy	Van den Bergh	Whitson	Soloway	Selvadurai	D'Amico	Entire cohort
Insignificant cancer, classical definition								
Correctly eligible for AS	3	9	3	9	8	12	9	20
Incorrectly eligible for AS	41	137	60	160	98	294	197	396
Total eligible for AS	44	146	63	169	106	306	206	415
Percent of eligible patients correctly identified as insignificant PC	7	6	5	5	8	4	4	5
Insignificant cancer, ERSPC definition								
Correctly eligible for AS	17	45	22	47	36	66	51	99
Incorrectly eligible for AS	27	101	41	122	70	240	155	316
Total eligible for AS	44	146	63	169	106	306	206	415
Percent of eligible patients correctly identified as insignificant PC	39	31	35	28	34	22	25	24
Organ-confined low-grade cancer								
Correctly eligible for AS	28	81	39	88	57	134	97	181
Incorrectly eligible for AS	16	65	24	81	49	172	109	234
Total eligible for AS	44	146	63	169	106	306	206	415
Percent of eligible patients correctly identified as insignificant PC	64	55	62	52	54	44	47	44

Abbreviations: AS = active surveillance; AUC = area under the receiver operator characteristic curve; ERSPC = European Randomised Study of screening for prostate cancer; PC = prostate cancer.

Table 6. Comparison of accuracy of tools in unscreened UK and screened USA data sets

Approach	Insignificant cancer, classical definition				Insignificant cancer, ERSPC definition				Organ-confined low-grade cancer			
	This study		Iremashvili <i>et al</i> , 2012		This study		Iremashvili <i>et al</i> , 2012		This study		Iremashvili <i>et al</i> , 2012	
	AUC	95% confidence interval	AUC	95% confidence interval	AUC	95% confidence interval	AUC	95% confidence interval	AUC	95% confidence interval	AUC	95% confidence interval
Tosoian	0.53	0.44–0.62	0.63	0.59–0.68	0.54	0.50–0.59	0.6	0.56–0.63	0.54	0.51–0.58	0.59	0.54–0.63
Adamy	0.58	0.45–0.70	0.63	0.58–0.67	0.57	0.51–0.63	0.62	0.57–0.67	0.59	0.54–0.63	0.59	0.54–0.64
Van den Bergh	0.51	0.42–0.60	0.68	0.63–0.72	0.55	0.50–0.59	0.64	0.60–0.69	0.56	0.52–0.59	0.62	0.57–0.67
Whitson	0.55	0.43–0.67	0.61	0.57–0.65	0.55	0.49–0.60	0.59	0.55–0.64	0.57	0.52–0.62	0.56	0.51–0.61
Soloway	0.6	0.48–0.72	0.68	0.64–0.73	0.57	0.52–0.63	0.63	0.58–0.67	0.55	0.51–0.60	0.6	0.55–0.65
Any criteria	0.42	0.30–0.53	0.61	0.57–0.65	0.45	0.40–0.50	0.58	0.54–0.63	0.5	0.46–0.53	0.55	0.50–0.60
All criteria	0.53	0.44–0.62	0.63	0.59–0.68	0.54	0.50–0.58	0.6	0.56–0.64	0.54	0.51–0.57	0.58	0.54–0.62

Abbreviations: AUC = area under the receiver operator characteristic curve; ERSPC = European Randomised Study of screening for prostate cancer.

Data from maturing AS cohorts suggest the short-term safety of an inclusive approach (Klotz, 2013; Selvadurai *et al*, 2013). However, 30% of patients on AS undergo radical treatment within 5 years. Upgrading and upstaging at surgery were common events in our cohort. It is possible that with increasing follow-up the number of patients failing conservative therapy will increase and while they are on AS they continue to age, rendering them less suitable for surgical treatment.

Considered together these data suggest a substantial underdiagnosis and would justify exhaustive investigation early in the course of an AS programme. The approach in our unit, in an attempt to minimise underdiagnosis, following a diagnosis of low-risk PC is to perform a multiparametric MRI with repeat TRUS at 3 months. The MRI is performed after a suitable period following

the diagnostic biopsy to exclude false positives due to biopsy artefact. If the MRI demonstrates a lesion likely to have been missed on TRUS-guided biopsy, targeted transperineal biopsy is performed. Inter-operator variability in interpreting MRI means that in some centres the use of systematic transperineal template biopsy is rational (Ayres *et al*, 2012). Where expertise in MRI is available, targeted biopsy may help rule out aggressive disease for those men who are unwilling to accept the inaccuracy associated with staging and grading of PC by TRUS biopsy, PSA and clinical examination.

This study highlights the uncertainty associated with selecting men for AS; although long-term data from AS series are awaited, it is clear that there is an urgent need for a biomarker by which to differentiate indolent from aggressive PC. Heterogeneity within PC

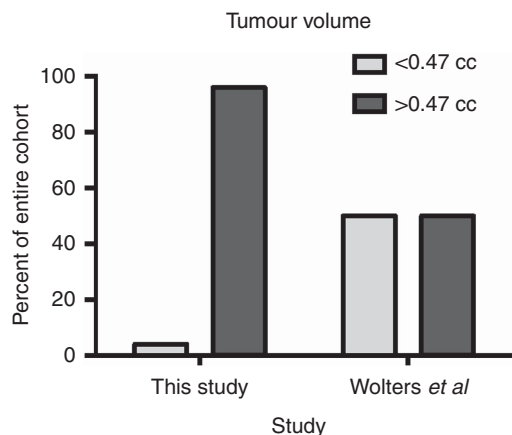


Figure 1. Bar chart of distribution of tumour volumes in screened and unscreened populations (we acknowledge technical differences in the method of calculation of tumour volume).

tumours means that the sampling error inherent to random or systemic needle biopsy of the prostate will continue to result in misdiagnosis. This error is unlikely to be negated using genomic or expressome analysis of tumour samples, where sampling error is still a problem. Further work should focus on imaging tests where areas suspicious for aggressive cancer can be identified and targeted for biopsy. Multiparametric magnetic resonance imaging is currently the best way to identify significant PC. In the future, positron emission tomography using tracers taken up specifically by aggressive tumour foci may prove useful.

CONCLUSIONS

Tools to identify insignificant PC are inaccurate in an unscreened population. This illustrates the great caution needed when using a tool developed and validated in one population in another population, particularly when PSA screening is a confounding factor. In counselling patients for AS, the surgeon should be explicit regarding uncertainty in predicting stage/grade despite apparent short-term safety. There is an urgent need for development of a means by which to exclude aggressive PC in patients wishing to undergo conservative treatment.

ACKNOWLEDGEMENTS

We are grateful to Praful Ravi, Hannah Rickman, Saurabh Singh and Robert Walsh, Medical Students at the University of Cambridge, for their help with data entry. We acknowledge the support of the National Institute for Health Research, which funds the Cambridge Bio-medical Research Centre, Cambridge, UK, and the support of the National Cancer Research ProMPT (Prostate Cancer: Mechanisms of Progression and Treatment) collaborative (grant code G0500966/75466), which has funded tissue and urine collections in Cambridge.

REFERENCES

Adamy A, Yee DS, Matsushita K, Maschino A, Cronin A, Vickers A, Guillonneau B, Scardino PT, Eastham JA (2011) Role of prostate specific antigen and immediate confirmatory biopsy in predicting progression during active surveillance for low risk prostate cancer. *J Urol* **185**: 477–482.

Albertsen PC, Hanley JA, Fine J (2005) 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* **293**: 2095–2101.

Andriole GL, Crawford ED, Grubb 3rd RL, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Clapp JD, Rathmell JM, Riley TL, Hayes RB, Kramer BS, Izmirlan G, Miller AB, Pinsky PF, Prorok PC, Gohagan JK, Berg CD. PLCO Project Team (2009) Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* **360**: 1310–1319.

Ayres BE, Montgomery BS, Barber NJ, Pereira N, Langley SE, Denham P, Bott SR (2012) The role of transperineal template prostate biopsies in restaging men with prostate cancer managed by active surveillance. *BJU Int* **109**: 1170–1176.

Bill-Axelsson A, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C, Nordling S, Häggman M, Andersson SO, Bratell S, Spångberg A, Palmgren J, Steineck G, Adami HO, Johansson JE. SPCG-4 Investigators (2011) Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* **364**: 1708–1717.

Carter HB, Albertsen PC, Barry MJ, Etzioni R, Freedland SJ, Greene KL, Holmberg L, Kantoff P, Konety BR, Murad MH, Penson DF, Zietman AL (2013) Early detection of prostate cancer: AUA Guideline. *J Urol* **190**: 419–426.

D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ, Wein A (1998) Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* **280**: 969–974.

DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**: 837–845.

Draisma G, Boer R, Otto SJ, van der Cruisjen IW, Damhuis RA, Schroder FH, de Koning HJ (2003) Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst* **95**: 868–878.

Graham J, Baker M, Macbeth F, Titshall V (2008) Diagnosis and treatment of prostate cancer: summary of NICE guidance. *BMJ* **336**: 610–612.

Iremashvili V, Pelaez L, Manoharan M, Jorda M, Rosenberg DL, Soloway MS (2012) Pathologic prostate cancer characteristics in patients eligible for active surveillance: a head-to-head comparison of contemporary protocols. *Eur Urol* **62**: 462–468.

Kim TH, Jeon HG, Choo SH, Jeong BC, Seo SI, Jeon SS, Choi HY, Lee HM (2013) Pathological upgrading and upstaging of patients eligible for active surveillance according to currently used protocols. *Int J Urol*; e-pub ahead of print 29 October 2013; doi:10.1111/iju.12326.

Klotz L (2013) Strengthening evidence for active surveillance for prostate cancer. *Eur Urol* **63**: 108–110.

MacMahon PJ, Kennedy AM, Murphy DT, Maher M, McNicholas MM (2009) Modified prostate volume algorithm improves transrectal US volume estimation in men presenting for prostate brachytherapy. *Radiology* **250**: 273–280.

Melia J, Moss S, Johns L (2004) Rates of prostate-specific antigen testing in general practice in England and Wales in asymptomatic and symptomatic patients: a cross-sectional study. *BJU Int* **94**: 51–56.

Moyer VA (2012) Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **157**: 120–134.

Ochiai A, Troncospo P, Chen ME, Lloreta J, Babaian RJ (2005) The relationship between tumor volume and the number of positive cores in men undergoing multisite extended biopsy: implication for expectant management. *J Urol* **174**: 2164–2168, discussion 2168.

Parker C, Muston D, Melia J, Moss S, Dearnaley D (2006) A model of the natural history of screen-detected prostate cancer, and the effect of radical treatment on overall survival. *Br J Cancer* **94**: 1361–1368.

Royal College of Pathologists (2009) Dataset for histopathology reports for prostatic carcinoma. *Standards and Datasets for Reporting Cancers*, 2nd edn, p 29. The Royal College of Pathologists: London.

Selvadurai ED, Singhera M, Thomas K, Mohammed K, Woode-Amisshah R, Horwich A, Huddart RA, Dearnaley DP, Parker CC (2013) Medium-term outcomes of active surveillance for localised prostate cancer. *Eur Urol* **64**: 981–987.

Soloway MS, Soloway CT, Eldefrawy A, Acosta K, Kava B, Manoharan M (2010) Careful selection and close monitoring of low-risk prostate cancer patients on active surveillance minimizes the need for treatment. *Eur Urol* **58**: 831–835.

Stamey TA, Freiha FS, McNeal JE, Redwine EA, Whittemore AS, Schmid HP (1993) Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. *Cancer* **71**: 933–938.

- Tosoian JJ, Trock BJ, Landis P, Feng Z, Epstein JI, Partin AW, Walsh PC, Carter HB (2011) Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *J Clin Oncol* **29**: 2185–2190.
- van den Bergh RC, Roemeling S, Roobol MJ, Aus G, Hugosson J, Rannikko AS, Tammela TL, Bangma CH, Schroder FH (2009) Outcomes of men with screen-detected prostate cancer eligible for active surveillance who were managed expectantly. *Eur Urol* **55**: 1–8.
- van der Kwast TH, Amin MB, Billis A, Epstein JI, Griffiths D, Humphrey PA, Montironi R, Wheeler TM, Srigley JR, Egevad L, Delahunt B (2011) International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 2: T2 substaging and prostate cancer volume. *Mod Pathol* **24**: 16–25.
- Whitson JM, Porten SP, Hilton JF, Cowan JE, Perez N, Cooperberg MR, Greene KL, Meng MV, Simko JP, Shinohara K, Carroll PR. The relationship between prostate specific antigen change and biopsy progression in patients on active surveillance for prostate cancer. *J Urol* (2011) **185**(5): 1656–1660.
- Wilt TJ, Brawer MK, Jones KM, Barry MJ, Aronson WJ, Fox S, Gingrich JR, Wei JT, Gilhooly P, Grob BM, Nsouli I, Iyer P, Cartagena R, Snider G, Roehrborn C, Sharifi R, Blank W, Pandya P, Andriole GL, Culkin D, Wheeler T. Prostate Cancer Intervention versus Observation Trial (PIVOT) Study Group (2012) Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med* **367**: 203–213.
- Wolters T, Roobol MJ, van Leeuwen PJ, van den Bergh RC, Hoedemaeker RF, van Leenders GJ, Schroder FH, van der Kwast TH (2011) A critical analysis of the tumor volume threshold for clinically insignificant prostate cancer using a data set of a randomized screening trial. *J Urol* **185**: 121–125.
- Xia J, Gulati R, Au M, Gore JL, Lin DW, Etzioni R (2013) Effects of screening on radical prostatectomy efficacy: the prostate cancer intervention versus observation trial. *J Natl Cancer Inst* **105**: 546–550.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.