International multicentre study examining selection criteria for active surveillance in men undergoing radical prostatectomy

L-M Wong*,1, DE Neal1, RB Johnston1, N Shah1, N Sharma1, AY Warren2, CM Hovens3, S Larry Goldenberg4, ME Gleave4, AJ Costello2 and NM Corcoran1,4

1Department of Urology, Addenbrooke’s Hospital, Cambridge, UK; 2Department of Pathology, Addenbrooke’s Hospital, Cambridge, UK; 3The Australian Prostate Cancer Centre at Epworth and Departments of Urology and Surgery, Royal Melbourne Hospital and University of Melbourne, Parkville, VIC, Australia; 4The Vancouver Prostate Centre, Department of Urological Sciences, University of British Columbia, Vancouver, BC, Canada

BACKGROUND: The controversies concerning possible overtreatment of prostate cancer, highlighted by debate over PSA screening, have highlighted active surveillance (AS) as an alternative management option for appropriate men. Regional differences in the underlying prevalence of PSA testing may alter the pre-test probability for high-risk disease, which can potentially interfere with the performance of selection criteria for AS. In a multicentre study from three different countries, we examine men who were initially suitable for AS according to the Toronto and Prostate Cancer Research International: Active Surveillance (PRIAS) criteria, that underwent radical prostatectomy (RP) in regards to:

1. the proportion of pathological reclassification (Gleason score ≥7, ≥pT3 disease),
2. predictors of high-risk disease,
3. create a predictive model to assist with selection of men suitable for AS.

METHODS: From three centres in the United Kingdom, Canada and Australia, data on men who underwent RP were retrospectively reviewed (n = 2329). Multivariable logistic regression was performed to identify predictors of high-risk disease. A nomogram was generated by logistic regression analysis, and performance characterised by receiver operating characteristic curves.

RESULTS: For men suitable for AS according to the Toronto (n = 800) and PRIAS (410) criteria, the rates for upgrading were 50.6, 42.7%, and upstaging 17.6, 12.4%, respectively. Significant predictors of high-risk disease were:

- Toronto criteria: increasing age, cT2 disease, centre of diagnosis and number of positive cores.
- PRIAS criteria: increasing PSA and cT2 disease.

Cambridge had a high pT3a rate (26 vs 12%). To assist selection of men in the United Kingdom for AS, from the Cambridge data, we generated a nomogram predicting high-risk features in patients who meet the Toronto criteria (AUC of 0.72).

CONCLUSION: The proportion of pathological reclassification in our cohort was higher than previously reported. Care must be used when applying the AS criteria generated from one population to another. With more stringent selection criteria, there is less reclassification but also fewer men who may benefit from AS.

British Journal of Cancer (2012) 107, 1467–1473. doi:10.1038/bjc.2012.400 www.bjcancer.com
Published online 4 October 2012
© 2012 Cancer Research UK

Keywords: prostate cancer; active surveillance; radical prostatectomy; nomogram; pathology; United Kingdom

Treatment paradigms for small-volume, low-grade prostate cancer (PCa) are currently moving away from radical treatments. The recent controversial US preventative task force report (Lin et al, 2011) giving PSA screening a ‘D’ rating, prompted by the results from the PLCO and ERSP trials, has highlighted the issues associated with overdiagnosis and overtreatment of PCa. In follow-up of over 3,500 patients across seven active surveillance (AS) case series, the cancer-specific survival for the cohort is 99.7% (Cooperberg et al, 2011), though median follow-up remains relatively short (2–7 years). Active surveillance has emerged as a viable management option and should be offered to patients with low-risk cancer (National Institute for Health and Clinical Excellence, 2008; National Institute of Health, 2011).

There are many different inclusion criteria for AS published in the literature. While all are variations on the model developed by Epstein et al (1994), the discrepancies between them reflect the uncertainty in appropriate cutoffs to distinguish indolent from high-risk cancer. In addition, regional differences in the underlying prevalence of PSA testing in the community alters the pre-test probability for high-risk disease (Moore et al, 2009) in defined ‘low-risk’ cohorts, which potentially interferes with the performance of selection rules when applied to populations distinct from which they were generated.

In our study, we examine application of AS selection rules to a combined Australian, British and Canadian group of patients who underwent radical prostatectomy (RP). Our primary objective was to document the proportion of pathological reclassification from prostate biopsy to RP. Secondary aims included analysis for predictors of high-risk disease and creation of a predictive model to assist with selection of men suitable for AS.
MATERIALS AND METHODS

Pooled prospectively collected data from Addenbrooke’s Hospital, Cambridge, United Kingdom (2005–2010); The Vancouver Prostate Centre, Canada (1995–2010); and the Australian Prostate Cancer Centre at Epworth, Melbourne, Australia (2003–2010) were retrospectively analysed. All patients had their RP specimens discussed at centralised multidisciplinary meetings and evaluated by dedicated genitourinary pathologists. Cambridge and Vancouver had routine centralised review of biopsies. Ethics approval in all three centres covers the use of collected clinical information for prognostic studies.

A summary of the literature review for published inclusion criteria used for AS is shown in Table 1 (Hardie et al., 2005; Warlick et al., 2006; Dall’Era et al., 2008; Klotz et al., 2009; van den Bergh et al., 2009; Soloway et al., 2010; Adamy et al., 2011). In selecting which AS criteria to apply to our series, we used the University of Toronto criteria, described by Klotz et al. (2009), from the first protocol-based AS prospective study and those published from the Prostate Cancer Research International: Active Surveillance (PRIAS) originating from the ERSPC (van den Bergh et al., 2009). The preoperative characteristics for these two groups are shown, overall and by treatment centre, in Tables 2a and 2b. The pathology results from RP, including proportions reclassified and final risk group, also divided by treatment centre, are shown in Tables 3a and b. Overall, for those satisfying the Toronto criteria, 50.6% were upgraded to GS ≥7 and 17.6% upstaged to pT3/4. The reclassification rates for the PRIAS criteria group were 40.5% and 12.4%, respectively. For both groups, the majority of GS upgrading consisted of 3+4 disease (Toronto = 84% and PRIAS = 79%). In Cambridge, there was a relatively high rate of pT3a disease in the Toronto criteria group (26%) that decreased with PRIAS criteria (14%). Melbourne had a relatively high rate of Gleason 7 disease in final pathology for both the Toronto and PRIAS criteria groups (61 and 52%).

Given that study periods at different centres were different, to account for changes in biopsy technique, pathological interpretation and treatment patterns with time, we repeated the analysis in a restricted cohort (year ≥2003 and total number of cores taken at biopsy ≥8) more reflective of contemporary practice. Results of this subanalysis for the Toronto and PRIAS inclusion criteria are shown in Table 4 and were similar to the initial overall analysis with similar proportions of GS upgrading (49.4% Toronto, 41% PRIAS) and pT3/4 upstaging (Toronto 17.9% and PRIAS 11.3%).

Using standard reported clinico-pathological variables, we were unable to derive a more accurate model that predicted indolent disease than the PRIAS criteria. However, we were able to identify significant predictors for high-risk disease (Table 5a) by multivariate logistic regression analysis. For patients meeting the Toronto criteria, increasing age, number of positive cores, the

---

Table 1: Published criteria for active surveillance

| Institution          | Principal investigator | Clinical stage | PSA (ng ml⁻¹) | Biopsy Gleason score | PSA density (ng ml⁻¹) | No. of +ve cores | % of single core |
|----------------------|------------------------|----------------|---------------|----------------------|-----------------------|-----------------|-----------------|
| Royal Marsden        | Parker                 | T1-T2          | ≤15           | ≤3+4                 | ≤50% total cores      | —               | —               |
| University of Toronto| Klotz                  | T1c/T2a        | ≤10           | ≤6                   | —                     | —               | 50%             |
| PRIAS                | Shroder                | T1c            | ≤10           | ≤6                   | ≤0.20                 | 2               | 2               |
| UCSF                 | Carroll                | T1-T2          | ≤10           | ≤6                   | —                     | 1/3             | 50%             |
| MSK                  | Eastham                | T1-T2a         | ≤10           | ≤6                   | ≤3                    | <2              | ≤30%            |
| University of Miami  | Soloway                | ≤T2            | ≤6            | ≤6                   | ≥6                    | 2               | ≤20%            |
| John Hopkins Medical | Carter                 | T1c            | —             | ≤3, no pattern Gleason | ≤0.15                | 4 or 5          | 50%             |

Abbreviations: MSK = Memorial Sloan Kettering; PRIAS = Prostate Cancer Research International: Active Surveillance; PSA = prostate-specific antigen; UCSF = University of California, San Francisco.

RESULTS

Of the 2329 patients that had RP, 800 patients met the Toronto criteria for AS, and this number was reduced to 410 patients when the stricter PRIAS criteria were applied. The preoperative characteristics for these two groups are shown, overall and by treatment centre, in Tables 2a and 2b. The pathology results from RP, including proportions reclassified and final risk group, also divided by treatment centre, are shown in Tables 3a and b. Overall, for those satisfying the Toronto criteria, 50.6% were upgraded to GS ≥7 and 17.6% upstaged to pT3/4. The reclassification rates for the PRIAS criteria group were 40.5% and 12.4%, respectively. For both groups, the majority of GS upgrading consisted of 3+4 disease (Toronto = 84% and PRIAS = 79%). In Cambridge, there was a relatively high rate of pT3a disease in the Toronto criteria group (26%) that decreased with PRIAS criteria (14%). Melbourne had a relatively high rate of Gleason 7 disease in final pathology for both the Toronto and PRIAS criteria groups (61 and 52%).

Given that study periods at different centres were different, to account for changes in biopsy technique, pathological interpretation and treatment patterns with time, we repeated the analysis in a restricted cohort (year ≥2003 and total number of cores taken at biopsy ≥8) more reflective of contemporary practice. Results of this subanalysis for the Toronto and PRIAS inclusion criteria are shown in Table 4 and were similar to the initial overall analysis with similar proportions of GS upgrading (49.4% Toronto, 41% PRIAS) and pT3/4 upstaging (Toronto 17.9% and PRIAS 11.3%).

Using standard reported clinico-pathological variables, we were unable to derive a more accurate model that predicted indolent disease than the PRIAS criteria. However, we were able to identify significant predictors for high-risk disease (Table 5a) by multivariate logistic regression analysis. For patients meeting the Toronto criteria, increasing age, number of positive cores, the
presence of palpable disease as well as the centre of diagnosis were all significant predictors of the presence of high-risk disease; whereas for patients meeting the more stringent PRIAS criteria, only increasing PSA and the presence of palpable disease were significant. Total number of cores taken was analysed and not predictive.

A number of subgroup analyses were undertaken to search for predictors of more advanced disease. Examining for cancer present in a biopsy core (percentage) as a predictor in only the Cambridge cohort did not yield a significant result. Repeating multivariate logistic regression analyses, limited by year of surgery (X2003) and number of biopsy cores taken (X8) for high-risk disease (pT3 or Gleason sum X8) or pT3 alone, demonstrated similar predictors to those found for the entire cohort. There were no significant predictors of primary Gleason pattern 4 (as opposed to Gleason sum 8) identified.

We were unable to generate a nomogram from the whole cohort data because of difficulties modelling sampling error. To account for sampling error, we used an index of prostate size to number of cores taken; however, this did not significantly improve the performance of the model. As the rate of pT3a was high in the Cambridge data, we performed logistic regression analysis to identify predictors of pT3a/GS8 disease specifically for the Cambridge cohort alone (Table 5b). When only Cambridge data were analysed for the nomogram, PSAD was more predictive than PSA. From this analysis, a nomogram (Figure 1a) was generated that predicts individual risk of high-risk disease in UK men who meet the Toronto criteria, based on age, PSA density, number of positive cores.

### Table 2a Preoperative characteristics for patients suitable for AS according to the Toronto criteria – by centre

|                     | Total (total RP) | Cambridge | Melbourne | Vancouver |
|---------------------|------------------|-----------|-----------|-----------|
| Years of data       | 1995–2010        | 2005–2010 | 2003–2010 | 1995–2010 |
| N (total RP)        | 2329             | 700       | 790       | 839       |
| N (total AS)        | 800 (34.3%)      | 267 (40%) | 187 (31%) | 190 (32%) |
| Age (years)         |                  |           |           |           |
| Median (IQR)        | 61 (56.7 – 65)   | 61 (39 – 73) | 60 (42 – 74) | 61 (43 – 79) |
| PSA (ng ml−1)       |                  |           |           |           |
| Median (IQR)        | 5.8 (4.7 – 7.4)  | 6.4 (0.5 – 10) | 5.5 (0.3 – 10) | 5.5 (0.5 – 10) |
| Clinical stage      |                  |           |           |           |
| cT1                 | 570 (71.3%)      | 206 (77%) | 149 (80%) | 109 (58%) |
| cT2a                | 230 (28.7%)      | 59 (23%)  | 38 (20%)  | 80 (42%)  |
| PSA density*        |                  |           |           |           |
| Median (IQR)        | 0.1 (0.086 – 0.151) | 0.108 (0.011 – 0.816) | 0.114 (0.014 – 0.315) | 0.11 (0.012 – 0.371) |
| Biopsy cores taken  |                  |           |           |           |
| Median (range)      | 10 (2 – 30)      | 12 (4 – 30) | 11 (4 – 30) | 8 (2 – 13) |
| Number of positive cores | 2 (1 – 4) | 2 (1 – 14) | 3 (1 – 12) | 2 (1 – 8) |

### Table 2b Preoperative characteristics for patients suitable for AS according to the PRIAS criteria – by centre

|                     | Total (total RP) | Cambridge | Melbourne | Vancouver |
|---------------------|------------------|-----------|-----------|-----------|
| N (total RP)        | 410 (18%)        | 134 (19%) | 114 (14%) | 162 (19%) |
| N (total AS)        | 103 (18%)        | 39 (9%)   | 31 (8%)   | 33 (5%)   |
| Age (years)         |                  |           |           |           |
| Median (IQR)        | 62 (56.3 – 65)   | 62 (45 – 70) | 59 (48 – 73) | 62 (43 – 73) |
| PSA (ng ml−1)       |                  |           |           |           |
| Median (IQR)        | 5.6 (4.3 – 7)    | 6.3 (2.7 – 10) | 5.6 (0.3 – 10) | 5.4 (1.5 – 0.4) |
| Clinical stage      |                  |           |           |           |
| cT1                 | 287 (70%)        | 109 (81%) | 91 (80%)  | 87 (54%)  |
| cT2a                | 123 (30%)        | 25 (19%)  | 23 (20%)  | 75 (46%)  |
| PSA density*        |                  |           |           |           |
| Median (IQR)        | 0.1 (0.071 – 0.13) | 0.99 (0.02 – 0.19) | 0.1 (0.01 – 0.19) | 0.1 (0.01 – 0.2) |
| Biopsy cores taken  |                  |           |           |           |
| Median (range)      | 9 (2 – 24 – IQR) | 12 (6 – 24) | 10 (3 – 17) | 8 (2 – 12) |
| Number of positive cores | 1 (1 – 2)   | 1 (1 – 2) | 2 (1 – 2) | 2 (1 – 2) |

Abbreviations: AS = active surveillance; IQR = interquartile range; PRIAS = Prostate Cancer Research International: Active Surveillance; PSA = prostate-specific antigen; RP = radical prostatectomy. *Not part of the original Klotz criteria but reported for comparison with Van den Bergh.
positive biopsy cores, and the presence or absence of palpable disease. The logistic regression equation for the nomogram is also included to facilitate future validation studies (Figure 1b).

Assessment of the nomogram to predict the presence of high-risk disease in a randomly selected evaluation cohort revealed reasonable accuracy (AUC of 0.72) (Figure 2). A calibration plot, comparing nomogram-predicted probabilities to actual proportions of high-risk disease, is shown in Figure 3. This shows that our nomogram underestimates the observed probability of high-risk disease. The two-component (calibration and AUC calculation) decomposition Brier score was 0.199. External validation for the nomogram generated from the Cambridge data, using Melbourne and Vancouver data, showed an AUC of 0.68 and 0.55, respectively.

**DISCUSSION**

It is clear that published inclusion criteria for AS (Table 1) vary in their stringency. The criteria at John Hopkins Medical Institute are the strictest but some centres elsewhere have accepted Gleason 7 (usually 3+4), PSA levels up to 15 ng ml⁻¹ and all clinical T2 disease. Furthermore, not all centres use PSAD (≤0.15–0.20), number of positive cores (≤2, ≤3 or 1/3 of total) and percentage (≤20–50%) of single core involvement to enrol patients. It has been demonstrated previously (Suardi et al, 2008; Conti et al, 2009; Mufarrij et al, 2010) that increases in stringency can decrease the rates of adverse pathological features but will also substantially decrease the number of men suitable for AS. Until the ProtecT trial reports (Lane et al, 2010), we will not know the criteria that predict the most important outcome of AS, namely death from PCA.

Our paper reports combined results from the Australian, British and Canadian academic centres. These results are compared with the American (UCSF, n = 331 (Conti et al, 2009)) and European (Milan and Hamburg, n = 2455 and Milan, n = 85 (Suardi et al, 2008; Suardi et al, 2010)) cohorts, where the Toronto and PRIAS inclusion criteria were also used to select patients from a RP database. Applying the Toronto criteria, we found a higher than previously reported rate of Gleason score upgrading (50.6%) compared with those previously reported by the Milan group (14–15%). There was less upstaging and upgrading when using the stricter PRIAS criteria; however, our reclassification rates are similar to previous reports, the majority of GS upgrading was 3 or 1/3 of total) and percentage (≤20–50%) of single core involvement to enrol patients. It has been demonstrated previously (Suardi et al, 2008; Conti et al, 2009; Mufarrij et al, 2010) that increases in stringency can decrease the rates of adverse pathological features but will also substantially decrease the number of men suitable for AS. Until the ProtecT trial reports (Lane et al, 2010), we will not know the criteria that predict the most important outcome of AS, namely death from PCA.

![Table 3a](image.png) Pathological results, with upgrading and upstaging rates highlighted, from radical prostatectomy for patients suitable for the Toronto AS criteria

| Total (Toronto) | Cambridge | Melbourne | Vancouver | P |
|----------------|-----------|-----------|-----------|---|
| n              | 800       | 280       | 248       | 272 |
| Pathological GS 6 | 395 (47.8%) | 157 (56%) | 94 (38%) | 144 (53%) | <0.001 |
| Upgrading 7 | 389 (48.6%) | 117 (43%) | 151 (61%) | 121 (44%) |
| 3+4 304 (87.4%) | 108 (92%) | 133 (88%) | 99 (82%) | 0.049 |
| 4+3 49 (12.6%) | 9 (8%) | 18 (12%) | 22 (18%) |
| 8–10 16 (2%) | 6 (2%) | 3 (1%) | 7 (3%) |
| pT pT2 659 (82.4%) | 205 (73%) | 216 (87%) | 238 (88%) | <0.001 |
| Upstaging pT3/4 | 141 (17.6%) | 75 (27%) | 32 (13%) | 34 (12%) |
| EPE 136 (17%) | 73 (26%) | 30 (12%) | 32 (12%) |
| SVI 8 (1%) | 2 (1%) | 3 (1%) | 3 (1%) |
| Margins Negative 675 (84.4%) | 239 (85%) | 217 (88%) | 219 (81%) | 0.077 |
| Positive 125 (15.6%) | 41 (15%) | 31 (12%) | 53 (19%) |
| Percent tumour (n = 716) | 5 | 5 | 3 | 10 | <0.001 |
| Median Range | 3–15 | 4–10 | 1.5–9 | 5–20 |
| Final risk group* Low 362 (45.3%) | 135 (44%) | 91 (37%) | 136 (50%) | <0.001 |
| Intermediate 286 (35.8%) | 65 (23%) | 123 (50%) | 98 (36%) |
| High 152 (19%) | 80 (29%) | 13 (13%) | 38 (14%) |
| CAPRA-S risk groups 0–2 629 (78.6%) | 208 (77.9%) | 152 (82.3%) | 143 (75.3%) | 0.42 |
| 3–5 165 (20.6%) | 58 (21.7%) | 34 (18.2%) | 44 (23.2%) |
| ≥6 6 (0.8%) | 1 (0.4%) | 1 (0.5%) | 3 (1.5%) |

**Table 3b** Pathological results, with upgrading and upstaging rates highlighted, from radical prostatectomy for patients suitable for the PRIAS criteria

| Total (PRIAS) | Cambridge | Melbourne | Vancouver | P |
|----------------|-----------|-----------|-----------|---|
| n              | 410       | 134       | 114       | 162 |
| Pathological GS 6 | 235 (57.3%) | 87 (65%) | 53 (46%) | 95 (58%) | 0.005 |
| Upgrading 7 | 166 (40.5%) | 44 (33%) | 59 (52%) | 63 (39%) |
| 3+4 138 (83.1%) | 40 (91%) | 52 (88%) | 46 (73%) | 0.023 |
| 4+3 28 (16.9%) | 4 (9%) | 7 (12%) | 17 (27%) |
| 8–10 9 (2.2%) | 3 (2%) | 2 (2%) | 4 (3%) |
| pT pT2 359 (87.6%) | 115 (86%) | 104 (91%) | 140 (86%) | 0.43 |
| Upstaging pT3/4 | 51 (12.4%) | 19 (14%) | 10 (9%) | 22 (14%) |
| EPE 48 (11.7%) | 19 (14%) | 9 (8%) | 20 (12%) | 0.27 |
| SVI 4 (1%) | 0 (0%) | 1 (1%) | 3 (2%) |
| Margins Negative 364 (88.8%) | 122 (91%) | 103 (90%) | 139 (86%) | 0.3 |
| Positive 46 (11.2%) | 12 (9%) | 11 (10%) | 23 (14%) |
| Percent tumour (n = 367) | 5 | 5 | 2 | 10 | <0.001 |
| Median Range | 2–10 | 0.2–80 | 0.3–37 | 1–70 |
| Final risk group* Low 219 (53.4%) | 79 (59%) | 52 (46%) | 88 (54%) | 0.014 |
| Intermediate 133 (32.4%) | 33 (25%) | 51 (45%) | 49 (30%) |
| High 53 (14.1%) | 22 (16%) | 11 (9%) | 25 (16%) |
| CAPRA-S risk groups 0–2 (low) 346 (84.9%) | 112 (89.6%) | 76 (87%) | 77 (79%) | 0.2 |
| 3–5 60 (14.6%) | 12 (9.6%) | 10 (12%) | 20 (20%) |
| ≥6 4 (1%) | 1 (0.8%) | 1 (1%) | 1 (1%) |

**Abbreviations:** AS = active surveillance; CAPRA = Cancer of the Prostate Risk Assessment Post-Surgical Score; GS = Gleason sum; PRIAS = Prostate Cancer Research International: Active Surveillance. *Risk group patterned on D’Amico system with pT and Gleason sum from radical prostatectomy (instead of cT and biopsy Gleason sum). Risk groups 0–2: low; 3–5: intermediate; ≥6: high.
The number of biopsy cores taken for each centre was: Cambridge (median = 12, IQR10-12), Melbourne (median = 10, IQR8-13) and Vancouver (median = 8, IQR8-8). The standard extended core template for prostate biopsy consists of 10–12 cores, and hence particularly in Vancouver biopsy under-sampling could affect our results. To minimise biopsy sampling error, Adamy et al (2011) suggest immediate confirmatory biopsy before commencing AS. There may be inter-observer variation influencing the Melbourne data as ~50% of biopsies are reported by community pathologists, whereas all RP specimens are read by two specialist uro-pathologists. In addition, most of the upgrading was predominantly Gleason 3+4 disease (84% of those upgraded (Suardi et al, 2008)). Results for carefully selected men on AS with intermediate-risk disease (Gleason sum 7 or CAPRA score 3–5) have been published suggesting that within limited follow-up (4years) outcomes were similar to men with GS 3+3 disease (Cooperberg et al, 2011).

### Table 4

| Clinical stage | Toronto criteria | PRIA criteria | P |
|----------------|------------------|--------------|---|
| Age (years)    | Medium (IQR)     | Median (IQR) |   |
| Median (IQR)   | 61 (57–65)       | 61 (57–65)   | 0.55 |
| PSA (ng/mL⁻¹)  | Median (IQR)     | 5.8 (4.7–7.4) | 0.018 |
| Clinical stage | cT1              | cT2a/cT2     |    |
| Median (range) | 467 (72.5%)      | 177 (27.5%)  | 0.45 |
| PSA density²   | Median (IQR)     | 0.111 (0.082–0.147) | 0.098 (0.07–0.127) | <0.001 |
| Biopsy cores taken | Median (range) | 10 (8–12) | 10 (8–12) |
| Number of positive cores | Median (IQR) | 2 (1–4) | 1 (1–2) | 0.48 |
| Centre (n, %)  | Cambridge        | Vancouver    | Melbourne |
| Number of positive cores | 267 (41.5%) | 190 (39%) | 187 (29%) |
| Number of positive cores | 125 (40.3%) | 87 (28.1%) | 98 (31.6%) |
| Prostatectomy Gleason score | ≤ 6 | 311 (48.3%) | 176 (56.8%) | 0.24 |
| 7              | 318 (49.4%)      | 127 (41%)    |   |
| 3 + 4          | 273 (85.8%)      | 102 (80.2%)  |   |
| 4 + 3          | 45 (14.2%)       | 25 (19.7%)   |   |
| 8–10           | 15 (2.3%)        | 7 (2.2%)     |   |
| pT              | pT2              | pT3/4        | 0.024 |
| Median (range) | 529 (82.1%)      | 115 (17.9%)  |   |
| pT3/4          | 275 (88.7%)      | 35 (11.3%)   |   |
| EPE             | No               | Yes          | 0.008 |
| Median (range) | 533 (82.8%)      | 277 (89.4%)  |   |
| EPE             | 111 (17.2%)      | 33 (10.6%)   |   |
| SV              | No               | Yes          | 0.87 |
| Median (range) | 637 (98.9%)      | 307 (99%)    |   |
| SV              | 7 (1.1%)         | 3 (1%)       |   |

### Table 5a

| Predictors of high-risk(+) disease for the Cambridge cohort, Toronto AS selection criteria, | OR 95% CIs | P |
|-----------------------------------------------------------|------------|---|
| Age                                                       | 1.04       | 1.01–1.07 | 0.02 |
| PSA                                                       | 1.08       | 0.98–1.19 | 0.1  |
| cT                                                        | 1.54       | 1.01–2.34 | 0.045 |
| Total no. of cores taken                                  | 0.94       | 0.88–1.01 | 0.085 |
| No. of positive cores                                     | 1.25       | 1.14–1.37 | <0.001 |
| Centre                                                    |            |            |     |
| Vancouver                                                | 1.03       | 0.59–1.79 | 0.91 |
| Cambridge                                                | 2.85       | 1.68–4.84 | 0.001 |
| Melbourne                                                | 1.03       | 0.59–1.79 | 0.91 |

### Table 5b

| Predictors of high-risk(+) disease for combined three centres, the Toronto and PRIAS selection criteria for AS groups | OR 95% CIs | P |
|-----------------------------------------------------------------------------------------------------------------|------------|---|
| Age                                                                | 1.06       | 1  | 0.13 |
| PSA                                                                | 1.71       | 1.03 | 0.037 |
| cT                                                                | 2.71       | 1.2  | 6.12 | 0.017 |
| Positive cores                                                     | 1.24       | 1.06 | 1.46 | 0.008 |

Abbreviations: AS = active surveillance; CI = confidence interval; OR = odds ratio; PRIAS = Prostate Cancer Research International: Active Surveillance; PSA = prostate-specific antigen; PSA density² = not part of original Klotz criteria but reported for comparison to Van den Bergh.

Cambridge had a high proportion of EPE in its Toronto criteria group (26% vs 12%, P <0.001) compared with the other two centres. As pathology for biopsy and RP is centrally reviewed in Cambridge and an extended template used for biopsy, this high rate of pT3a disease could also be attributable to a low uptake of PSA testing in the United Kingdom. Melia (2005) compared worldwide rates of PSA testing in 2005 and, allowing for lack of standardized data, found rates of PSA testing in the United Kingdom considerably lower than elsewhere. At a similar period of time (year 2000), the rate of PSA testing in the United Kingdom was 5.4 tests per 100 men per annum (men 45–84 years old) compared with the United States, 38% of black men and 31% of white men (≥65 years); Italy, 26.9% men (≥40 years); Australia, 23% men (40–70 years); and Canada (Beaulac et al, 2006), 47.5% (≥50 years).

The few previously published predictive tools for AS selection have focused on calculating likelihood of indolent, low-volume/low-grade or insignificant PCa rather than high-risk features (Kattan et al, 2003; Ochiai et al, 2005; Nakanishi et al, 2007; Steyerberg et al, 2007; O’Brien et al, 2011). Nakanishi et al’s nomogram, using a cohort of 258 men from the Canada and the United States, is specific for men with a single-positive biopsy core and uses age, PSAD and tumour length in a core to predict indolent cancer (Nakanishi et al, 2007). Possibly having a single positive core is too stringent a criterion for a programme of AS. Kattan et al’s nomogram to predict small, moderately differentiated, confined tumours (Kattan et al, 2003) that was recently validated and updated (Steyerberg et al, 2007); however, its application is questionable as its data are based on sextant biopsies with most centres now performing 10–12 core biopsies. A nomogram generated from an Australian cohort was recently published (O’Brien et al, 2011), and for multiple probability cutoffs predicting indolent PCa, they gave coexisting rates of high-risk disease. It is likely that in the future men will be selected for AS programmes only if they have an mp-MRI that does not demonstrate any other unsuspected cancers, possibly coupled with a limited template biopsy. The result of the PIVOT trial (Wilt...
There is evidence that risk calculators are best applied to the Melbourne (AUC 0.68) and Vancouver (0.55) data to externally validate it. The nomogram performed poorer when we used the Melbourne disease for nearly all predicted values. Unsurprisingly, the nomogram consistently underestimates the observed proportion of high-risk disease (0.199). However, the calibration plot suggests that our nomogram was good (AUC 0.72 and two-component Brier score 0.163). Overall assessment of the performance of our nomogram to predict presence of high-risk disease in patients who satisfy the Toronto entry criteria for AS. Overall assessment of the performance of our nomogram was good (AUC 0.72 and two-component Brier score 0.199). However, the calibration plot suggests that our nomogram consistently underestimates the observed proportion of high-risk disease for nearly all predicted values. Unsurprisingly, the nomogram performed poorer when we used the Melbourne (AUC 0.68) and Vancouver (0.55) data to externally validate it. There is evidence that risk calculators are best applied to the population from which they are generated (Bhojani et al, 2009). An ideal nomogram would also include information on previous biopsies, family history of significant CaP and results of MRI imaging; however, this was not present in our data set. Given the low rate of PSA testing in the United Kingdom and the high rate of upstaging in our Cambridge cohort, our nomogram would be a reasonable tool for counselling UK patients in regards to AS.

Our study has limitations. Being retrospective, data collected were reliant on individual centres’ protocols and they did not include all information on tumour volume. The use of additional criteria such as length or percentage of a single core involved might reduce the amount of reclassification, but would likely reduce the number of patients to whom AS could be offered. Data on lymph node status were lacking as lymph node dissection for low-risk disease was according to surgeon preference, and not consistently performed or recorded. Lack of follow-up data for cancer recurrence or death is significant, as having pathological features of advanced disease on biopsy does not necessarily translate to poorer outcomes after surgery. Using a surgical cohort includes unforeseen biases in patient selection not addressed in AS criteria, such as age, comorbidities, family history, patient anxiety for intervention, and findings on imaging or institutional bias towards type of treatment. The median age of our cohort (61 years) is younger than that reported by AS cohorts (65 years; Carter et al, 2007; van den Bergh et al, 2009). Being a multicentre study, multiple pathologists reported biopsies and specimens, and the effect of inter-observer variation was not calculated. We also accept that, given that our study spanned a broad period of time, interpretive changes in pathological grading and clinical staging of PCa did occur (Thompson et al, 2005). Multiple surgeons performed the surgery though positive margin rates that were similar.

CONCLUSIONS

Our study examined the rates of reclassification of men from Australia, Britain and Canada who underwent RP who initially would have been suitable for AS, as defined by the Toronto and PRIAS criteria. Compared with previously reported cohorts from Europe and North America using the same AS selection criteria, we found significantly higher rates of upgrading and upstaging. Care must be used when applying AS criteria generated from one population to another distinct population. There is an onus on larger centres in individual countries to assess the performance of different criteria on their population before implementation in routine AS programmes and generate predictive tools from their own data sets. Use of increasingly stringent selection criteria may reduce reclassification but this must be balanced against the exclusion of a significant number of men from AS who may benefit from such an approach.

Figure 1 Nomogram to predict individual probability of high-risk PCa in UK men who meet the Klotz criteria. For a patient, each criteria (age, PSA density, number of positive cores and cT stage) translates to a number of points – read off the top line. The total number of points then corresponds to a probability of finding high-risk PCa if RP is performed. Logistic regression equation for nomogram: $\text{Log}(p/1-p) = -7.051 + 0.059(\text{age in years}) + 0.537(\text{PSAD in units of } 0.1 \text{ ng ml}^{-1}) + 0.996(\text{cT}) + 0.218(\text{number of positive cores})$.

Figure 2 Receiver operating characteristic curve to assess performance of nomogram predicting probability of high-risk PCa in UK men satisfying Klotz AS criteria.

Figure 3 Calibration plot comparing nomogram-predicted probabilities to actual proportions of high-risk disease.
ACKNOWLEDGEMENTS

We would like to thank Dr Karan Wadhwa for assistance with data entry in Cambridge.

REFERENCES

Adany 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(2): 477–482

Beaulac JA, Fry RN, Onysko J (2006) Lifetime and recent prostate specific antigen (PSA) screening of men for prostate cancer in Canada. Can J Public Health 97(3): 171–176

Bhojani N, Salomon L, Capitanio U, Suardi N, Shahrokh FS, Jeldres C, Zini L, Pharan D, Ois Pe Loquin O, Arjane P, Abbou C, De La Taille A, Montorsi F, Karakiewicz P (2009) External validation of the updated Partin tables in a Cohort of French and Italian men. Int J Radiat Oncol Biol Phys 73(2): 347–355

Carter HB, Ketteringham A, Warlick C, Metter EJ, Landis P, Walah PC (2007) Expectant management of prostate cancer with curative intent: an update of the Johns Hopkins experience. J Urol 178: 2359

Conti SL, Dall’Era M, Fradet V, Cowan JE, Simko J, Carroll PR (2009) Pathological outcomes of candidates for active surveillance of prostate cancer. J Urol 181: 1628–1634

Cooperberg MR, Cowan JE, Hilton JF, Reese AC, Zaid H, Porten SP, Shinohara K, Meng VM, Greene KL, Carroll PR (2011) Outcomes of active surveillance for men with intermediate-risk prostate cancer. JCO 29(2): 228–234

Cooperberg MR, Hilton JF, Carroll PR (2011) The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. Cancer 117: 5039–5046

Dall’Era MA, Komety BR, Cowan JE, Shinohara K, Stauf F, Cooperberg MR, Meng VM, Kane CJ, Perez N, Master VA, Carroll PR (2008) Active surveillance for the management of prostate cancer in a contemporary cohort. Cancer 112: 2664

Eng J (2006) (Updated 17 May 2006 (accessed 14 February 2011)) ROC Analysis: Web-based Calculator for ROC Curves. Johns Hopkins University: Baltimore. Available from: http://www.jrocfit.org

Epstein JI, Walsh PC, Carmichael M, Brendeler CB (1994) Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. JAMA 271: 368

Hardie C, Barker C, Norman A, Eales R, Horwich A, Huddart R, Deanrley D (2005) Early outcomes of active surveillance for localized prostate cancer. BJU Int 95: 956–960

Kattan MW, Eastham JA, Wheeler TM, Maru N, Scardino PT, Erbersdobler A, Graeven M, Huland H, Koh H, Shariat SF, Slawin KM, Ohori M (2003) Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors. J Urol 170: 1792–1797

Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A (2009) Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. J Clin Oncol 28: 126–131

Lane JA, Hamdy FC, Martin RM, Turner EL, Neal DE, Donovan JL (2010) Latest results from the UK trials evaluating prostate cancer screening and treatment: the CAP and ProteCt studies. Eur J Cancer 46(17): 3095–3101

Lin K, Crosswell JM, Koenig H, Lam C, Maltz A (2011) Prostate-Specific Antigen-Based Screening for Prostate Cancer: An Evidence Update for the U.S. Preventive Services Task Force [Internet]. Evidence syntheses, No. 90, Agency for Healthcare Research and Quality: Rockville, MD.

Available from: http://www.ncbi.nlm.nih.gov/books/NBK82303/

Mela J (2005) Part I: The burden of prostate cancer, its natural history, information on the outcome of screening and estimates of ad hoc screening with particular reference to England and Wales. BJU Int 98(Suppl 3): 4–15

Moore AL, Dimitropoulopou L, Lane A, Powell PH, Greenberg DC, Brown CH, Donovan JL, Hamdy FC, Martin RM, Neal DE (2009) Population-based prostate-specific antigen testing in the UK leads to a stage migration of prostate cancer. BJU Int 104(11): 1592–1598

Mufarrij P, Sankin A, Godoy G, Lepor H (2010) Pathologic outcomes of candidates for active surveillance undergoing radical prostatectomy. Urology 76(3): 689–692

Nakanishi H, Wang X, Ochiah A, Trpkov K, Yilmaz A, Donnelly JB, Davis JW, Troncos P, Babaian RJ (2007) A nomogram for predicting low-volume/low-grade prostate cancer. A tool for selecting patients for active surveillance. Cancer 110(11): 2441–2447

National Institute of Health (2011) Role of active surveillance in the management of men with localized prostate cancer. Draft Statement. National Institutes of Health State-of-the-Science Conference: Role of Active Surveillance in the Management of Men With Localized Prostate Cancer Convened 7 December 2009.

National Institute for Health and Clinical Excellence (2008) NICE Prostate Cancer: Diagnosis and Treatment. A Systematic Review (cited 13 May 2009). Available from http://www.nice.org.uk/nicemedia/pdf/CG58NICE_Guideline.pdf

O’Brien BA, Cohen RJ, Ryan A, Sengupta S, Mills J (2011) A new preoperative nomogram to predict minimal prostate cancer: accuracy and error rates compared to other tools to select patients for active surveillance. J Urol 186: 1811–1817

Ochiah A, Troncoso 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

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

Steyerberg EW, Roobol MJ, Kattan TH, van der Kwast TH, Koning HJ, Schroder FH (2007) Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. J Urol 177: 107–112

Suardi N, Briganti A, Gallina A, Salonia A, Karakiewicz PI, Capitanio U, Freschi M, Cesari A, Guazzoni G, Rigatti P, Montorsi F (2010) Testing the most stringent criteria for selection of candidates for active surveillance in patients with low-risk prostate cancer. BJU Int 105(11): 1548–1552

Suardi N, Capitanio U, Chun FK, Graeven M, Perrotte P, Schlommm T, Hasse A, Huland H, Erbersdobler A, Montorsi F, Karakiewicz PI (2008) Currently used criteria for active surveillance in men with low-risk prostate cancer. An analysis of pathologic features. Cancer 113(8): 2068–2072

Thompson IM, Canby-Hagino E, Lucia MS (2005) Stage migration and grade inflation in prostate cancer: Will Rogers meets Garrison Keilor. J Natl Cancer Inst 97: 1236–1237

den van Bergh RCN, Roemeling S, Roobol MJ, Roobol W, Schroder FH, Bangma CH (2007) Prospective validation of active surveillance in prostate cancer: the PRIAS study. Eur Urol 52: 1560–1563

den van 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

Warlick C, Trock BJ, Landis P, Epstein JI, Carter HB (2006) Delayed versus immediate surgical intervention and prostate cancer outcome. J Natl Cancer Inst 98: 355–357

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, Andreole CL, Culkin D. Wheeler for the Prostate Cancer Intervention versus Observation Trial (PIVOT) Study Group (2012) Radical prostatectomy versus observation for localized prostate cancer. NEJM 367(3): 203–213

This work is published under the standard license to publish, 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.