BJUI SUPPLEMENTS

Comparative analysis of three risk assessment tools in Australian patients with prostate cancer

David J. Tamblyn[1,2], Samarth Chopra[3,4], Changhong Yu[5], Michael W. Kattan[5], Carole Pinnock[1,6,7] and Tina Kopsaftis[1,7]

[1]Urology Unit, Division of Surgery, Repatriation General Hospital, Daw Park, SA, Australia, [2]School of Population Health and Clinical Practice, University of Adelaide, Australia, [3]St Vincent's Hospital Darlinghurst, NSW, Australia, [4]Garvan Institute, Darlinghurst, NSW, Australia, [5]Department of Quantitative Health Sciences, Cleveland Clinic, OH, USA, [6]School of Medicine, Flinders University of South Australia, Bedford Park, Australia, and [7]South Australian Prostate Cancer Clinical Outcomes Collaborative

SUMMARY

In many countries, prognostic tools, which draw on the experience of thousands of patients with cancer, are used to predict cancer outcomes, but accuracy varies. This paper compares the accuracy of three widely used tools predicting prostate cancer recurrence after surgery in Australian patients. The results show that all tools were good at predicting which patients were most likely to experience recurrence and which were least. However, prediction of absolute risk varied and the oldest tool was the most accurate.

What's known on the subject? and What does the study add?
Prognostic tools, such as the Cancer of the Prostate Risk Assessment (CAPRA) score and the 1998 Kattan and 2006 Stephenson nomograms, predicting biochemical recurrence after radical prostatectomy are widely used for treatment decision making and counselling patients. However, tools derived in certain cohorts tend to perform less well when they are applied to populations that are dissimilar in terms of population or disease characteristics, health systems or treatment practices. Some of the loss in accuracy of a prognostic tool is a consequence of unknown factors and hence the performance of a tool when applied to a different population is unknown and largely unpredictable.

This study validates these widely used tools in South Australian patients treated at three public hospitals. All three tools discriminated well according to risk of recurrence in these patients. However, when compared against observed rates of recurrence, it was found that predictions of recurrence varied widely between the three tools, suggesting that their use in counselling patients on such risk may not be appropriate. Interestingly, the oldest of the three tools (Kattan 1998) was the best predictor of absolute risk of recurrence. In the paper, this is linked to later adoption of updated Gleason grading, among other factors.

OBJECTIVE

- To compare performance of the CAPRA score and two commonly used risk assessment nomograms, the 1998 Kattan and the 2006 Stephenson, in an untested Australian cohort.

PATIENTS AND METHODS

- We present data on 635 men from the South Australian Prostate Cancer Clinical Outcomes Database who underwent radical prostatectomy between January 1996 and May 2009 and had all required variables for predicting biochemical recurrence (BCR).
- BCR was defined as prostate-specific antigen ≥0.2 ng/mL or secondary treatment for a rising prostate-specific antigen.
- Accuracy was evaluated using Harrell's concordance index, plotting calibration curves, and constructing decision analysis curves.

RESULTS

- Concordance indices were high for all three tools: 0.791, 0.787 and 0.744 for the 2006 Stephenson nomogram, CAPRA score and 1998 Kattan nomogram respectively.
- At 3 years, calibration of the tools (agreement between predicted and observed BCR-free probability) was close to ideal for the 1998 Kattan nomogram, whereas the 2006 Stephenson model underestimated and the CAPRA model overestimated BCR-free probability.
- The 1998 Kattan and 2005 CAPRA tools performed better than the 2006 Stephenson nomogram across a wide range of threshold probabilities using decision curve analysis.

CONCLUSION

- All three tools discriminate between patients' risk effectively.
- Absolute estimates of risk are likely to vary widely between tools, however, suggesting that models should be validated and, if necessary, recalibrated in the population to which they will be applied.
- Recent development does not mean a nomogram is more accurate for use in a particular population.

KEYWORDS

nomogram, prognosis, prostate cancer, prostate-specific antigen

INTRODUCTION

Prostate cancer is the most commonly diagnosed cancer in Australia, with an incidence now exceeding that of breast cancer [1]. Internationally, prognostic models predicting recurrence after definitive surgery for prostate cancer have become useful tools for assessing risk [2], treatment planning and patient counselling [3,4]. However most have not been validated for use in Australia and attitudes vary regarding their usefulness [5].

Three such tools which predict biochemical-recurrence-free probability (BCRFP) following radical prostatectomy (RP) and are in common international use today are the University of California and San Francisco Cancer of the Prostate Risk Assessment (CAPRA) score [6], the Kattan *et al.* nomogram (1998) [7] and the Stephenson *et al.* nomogram (2006) [8]. While the 1998 and 2006 nomograms as well as the CAPRA score have been successfully validated in the USA, European and Asian populations [9–17], their performance has not been tested solely in an Australian cohort. For effective comparison, candidate tools should be validated in the same cohort [18]. We therefore undertook a head-to-head analysis of the three risk assessment tools within a population of 635 men treated with RP between 1998 and 2009 to determine whether they are generalizable, or transportable, to a multi-institutional South Australian cohort [2,19,20].

PATIENTS AND METHODS

PATIENTS INCLUDED IN THE VALIDATION

The South Australian Prostate Cancer Clinical Outcomes Database (PCCOD) established in 1998 records clinical outcomes from patients with prostate cancer diagnosed and treated at the three major public hospitals in South Australia and some private practices, representing approximately 25% of all South Australian diagnoses. The patient characteristics in the database have been shown to be broadly representative of all South Australian prostate cancer diagnoses [21]. Data are collected independently of clinicians using electronic pathology databases and clinical case note reviews [21]. We prospectively collected data on 1376 men diagnosed with prostate cancer and treated with RP. Where clinical stage could not be decided a missing value was assigned.

Biochemical recurrence (BCR) was defined as a single PSA >0.2 ng/mL or a value of 0.2 ng/mL given a subsequent increase over 0.2 ng/mL. Secondary therapy following RP was deemed a recurrence if prescribed due to a rising PSA. Patients receiving secondary therapy with no rising PSA following RP were excluded from analysis ($N = 28$). Follow-up was measured until BCR, death or the most recent PSA value. The 1992 American Joint Committee on Cancer TNM classification was used for assignment of clinical stage. Predictions of BCRFP were generated using the nomogram equations by one of the authors (CY). The predicted probabilities of BCR for the CAPRA score were taken as quoted in the derivation paper [6]. For the CAPRA score, patients with a baseline PSA value <2 ng/mL were scored zero for their PSA component, as has been done in a previous validation [12].

STATISTICAL EVALUATION

The performance of the risk assessment tools was measured by comparing the concordance (calculated using Harrell's *C* index) and calibration of the tools. Harrell's *C* index is similar to the area under the curve statistic for receiver–operating characteristic plots but allows calculation of concordance in continuous and censored data (such as time to event data). Harrell's *C* measures the concordance between the predicted failure order of a pair of subjects and the observed order. Because 95% confidence intervals for Harrell's *C* statistic cannot be used to gauge whether one tool outperforms another, a test of the difference of Harrell's *C* between pairs of risk assessment tools was performed, as described by Newson [22], using the *lincom* command in STATA. While concordance is a measure of how well a tool can determine the relative risk of individual patients (compared with each other) in the population sample, calibration reflects how well the tool predicts an *absolute* outcome, such as the likelihood of recurrence at 3 years. Kaplan–Meier analysis was used to determine the actuarial 3-year BCRFP and this was plotted against predicted BCRFP to visually assess the calibration of the tools. Patients were either grouped by the seven CAPRA categories or split into quintiles ordered by predicted risk for each of the nomograms. This method of grouping is imperfect resulting in different sized groups between the tools and makes comparisons of the CAPRA score and the nomograms difficult. However, no alternative methods allowing a more robust comparison were devised and therefore comparisons of calibration plots must be approached with this in mind.

Finally, to directly compare the net benefit derived from each prediction tool, we implemented a decision curve analysis as previously described by Vickers and Elkin [23]. Decision curve analysis to compare prognostic models is not straightforward. Unlike models designed to inform a specific decision, it is unclear what actions will be informed by the models predicting BCR. A higher predicted likelihood of BCR may result in more frequent postoperative PSA tests; the offer of adjuvant radiotherapy; and/or a change in surgical technique. Each of these decisions may be made at different threshold probabilities of recurrence, and therefore the decision curve analysis must be interpreted carefully.

Baseline cohort characteristics were compared using the chi-squared test. Statistical significance was set at $P < 0.05$. All analyses were conducted using SPSS for Windows® v17 and STATA® v10. Ethics approval was obtained from the Repatriation General Hospital Ethical Review Committee (Protocol 122/10).

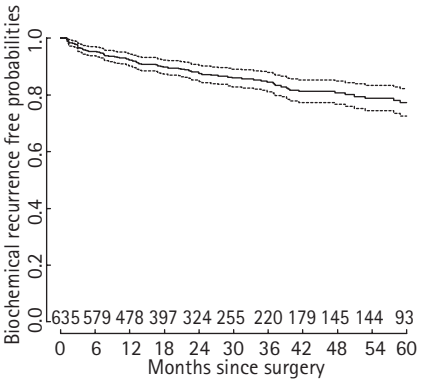
RESULTS

In all, 1376 patients received an RP between 1998 and June 2009 following diagnosis with transrectal ultrasound guided biopsy. Men with fewer than six cores taken at biopsy ($n = 25$), or who received neo-adjuvant therapy or adjuvant therapy following treatment in the absence of a rising PSA ($n = 28$), or who were missing variables required to calculate any of the risk tools ($n = 612$) were excluded from the analysis. Nearly half of the men excluded due to inadequate data were missing only clinical stage ($n = 301$) which is systemically poorly reported or reported in insufficient detail (i.e. T2 rather than T2a, b or c) in patient case notes. Patients with missing clinical stage tended to have had treatment

more recently, although they did not appear to be different from the patients included in the study. A further 76 men were excluded for having less than 3 months' follow-up.

In total, 635 patients were included in this study. Median age (range) was 62 (37–75) years and median time to event or censor was 2.8 (0.1–12) years. Freedom from BCR for the entire cohort was 84.5% and 77.3% at 3 and 5 years, respectively (Fig. 1). Baseline patient characteristics in the derivation data sets are presented in Table 1 and show broad overlap.

FIG. 1. Overall Kaplan–Meier estimated survival for PCCOD study cohort with 95% confidence intervals. Numbers above the x axis are the number of patients at risk at each time point.



Harrell's *C* concordance indices of all three risk assessment tools ranged from 0.744 to 0.791 (Table 2). Harrell's *C* concordance indices for the CAPRA and the Stephenson 2006 prediction tools were significantly better than Kattan 1998 ($P = 0.007$ and $P = 0.012$, respectively).

Calibration plots showing how closely the predicted BCRFP at 3 years matched the actual value for each nomogram and CAPRA are shown in Fig. 2. The best predictive tool in terms of calibration was the Kattan 1998 nomogram, where the agreement between predicted and observed BCRFP was close to ideal. The CAPRA tended to underestimate BCRFP in lower risk categories, and had wide confidence intervals depicting uncertainty and small numbers in higher risk categories. The Stephenson 2006 nomogram tended to overestimate BCRFP.

DISCUSSION

Despite their widespread use overseas, sophisticated risk assessment tools are not widely used in Australia and treatment decision making is more likely to be based on clinical opinion. To some extent this may be justified, because risk assessment tools in common use currently have been developed and validated overseas. The need for statistically based tools to be validated has been repeatedly stated [2,19,20].

The ability of the prediction tools to rank individuals in terms of their BCRFP, represented by Harrell's *C* coefficient, showed that all three tools performed well, varying from 0.744 (Kattan nomogram [7]) to 0.791 (Stephenson nomogram [8]). Nevertheless, actuarial BCRFP in our cohort (stratified by CAPRA score or by quintiles of risk for the two nomograms) was poorly predicted by both the CAPRA score and, in particular, the Stephenson 2006 nomogram. In contrast, the Kattan 1998 nomogram, whilst it did not rank patients quite as well in terms of risk, predicted absolute outcomes accurately.

TABLE 1 Socio-demographic and clinical variables of PCCOD and nomogram development data sets

Variables	PCCOD	Kattan 1998	Stephenson 2006	CAPRA 2005
No. of patients	635	983	1978	1439
Median age at diagnosis, years (range)	62 (37–75)	63 (31–81)		62 (NR)
Year of surgery	1996–2009	1983–1996	1987–2003	1991–2003
PSA (ng/mL), *n* (%)[a]				
≤10	478 (75.3)	689 (70.1)		1174 (81)
10 to ≤20	134 (21.1)	187 (19.0)		209 (15)
>20	23 (3.6)	107 (10.9)		20 (1)
median (IQR)	7.3 (5.21,10)		6.1 (4.4, 9.0)	
Clinical stage, *n* (%)[b]				
T1–T2	625 (98.4)	925 (94.1)	1891 (96)	1410 (98)
T3	10 (1.6)	58 (5.9)	88 (4)	29 (2)
Biopsy Gleason sum, *n* (%)[c]				
2–6	385 (60.6)	671 (68.3)[d]	1348 (68)	1068 (74)
7	200 (31.5)	213 (21.7)[d]	527 (27)	239 (17)[d]
8–10	50 (7.9)	99 (10.1)[d]	104 (5)	132 (9)[d]
Median follow-up,[e] years (IQR)	2.1 (1.1, 4.0)	2.5 (NR)	4.5 (2.5, 8)	2 (NR)
PSA threshold for recurrence (ng/mL)	0.2[f]	0.4[f]	0.4[f]	0.2[f]
Biochemically recurred, *n* (%)	100 (15.7)	196 (19.9)	220 (11.1)	210 (15)

IQR, interquartile range.

[a]PSA distribution in PCCOD was different from Kattan 1998, P < 0.001, and CAPRA 2005, P < 0.001. [b]The clinical stage distribution in PCCOD was different from Kattan 1998, P < 0.001, and Stephenson 2006, P = 0.001. [c]Gleason sum distribution in PCCOD was different from Kattan 1998, P < 0.001, Stephenson 2006, P = 0.001, and CAPRA 2005, P < 0.001. [d]Biopsy Gleason grades 1–3 + 4–5 have been grouped with Gleason 7 and grades 4–5 + 1–5 have been grouped with Gleason 8–10. [e]Median follow-up is calculated on patients who have not recurred; NR, not reported. [f]Require second value equal to or higher.

Prediction of absolute outcomes such as likelihood of recurrence at 3 years is important for counselling patients and decision making for clinicians. This is the primary function of both the Kattan and Stephenson nomograms. The equations required to predict BCRFP for the two nomograms are not readily available and their use for risk stratification in large analyses is therefore unlikely. On the other hand, while CAPRA can generate individual estimates of absolute risk, it is a particularly useful tool for risk stratification in research studies due to the ease of calculating the score for large cohorts.

Men of the PCCOD data set whose risk is deemed to be in the top quintile (highest) according to the Stephenson nomogram were predicted to have an average 3-year BCRFP of 83.5%. However, their actuarial BCRFP was closer to 60% (95% CI 49.9, 69.8). Such a difference is likely to be of great significance to a patient. The CAPRA score, on the other hand, was shown to underestimate BCRFP. Only 46% of men with CAPRA scores of 6 were predicted to be BCR-free at 3 years while the actuarial BCRFP of such men in the PCCOD cohort was nearer to 63%. This again is a difference likely to be important to a patient.

Surprisingly, the earlier developed Kattan *et al.* nomogram [7] was shown to be almost ideal in terms of absolute prediction of 3-year BCRFP (Fig. 2). When stratified into risk quintiles generated by the nomogram, the average actuarial BCRFP differed from the predicted BCRFP by only 0.2–3.1%. The confidence intervals of actuarial recurrence within the risk stratified groups were narrow and similar across all prediction tools, with the exception of the higher risk CAPRA categories which contained small numbers of patients.

The decision curve analysis (Fig. 3) reveals that for patients with a probability of 3-year BCR up to about 18%, both the Kattan 1998 and CAPRA 2005 prediction tools perform similarly. For probabilities between 18% and about 40%, the CAPRA 2005 prediction tool appears to perform better. If most decisions regarding future management (i.e. further or adjuvant interventions) are likely to occur in patients with higher likelihoods of BCR, then the predictions of the CAPRA 2005 tool will result in a higher net benefit. Both the CAPRA 2005 and Kattan 1998 tools appear to outperform the Stephenson 2006 nomogram.

There have been numerous validations of the three prediction tools. Graefen *et al.* [10] applied the Kattan 1998 nomogram to data of 6232 patients from the USA and three international sites. They reported a Harrell's *C* of 0.75 and described close predicted and actual rates of recurrence. These findings are strikingly similar to the performance of the Kattan 1998 nomogram in the PCCOD data set.

Lughezzani *et al.*'s study [11] compared the CAPRA and Stephenson prediction tools using data of 1976 patients treated between 1992 and 2006. The Stephenson nomogram in their study tends to overestimate BCRFP in low risk patients, and while confidence intervals for high risk patients include an ideal prediction, they are particularly wide for this range.

A concerning finding is that the Stephenson nomogram and the CAPRA score, arguably the most commonly used of the three prediction tools across the world, did not calibrate well in this cohort. There can be many causes for a prediction tool to underperform when applied to a new population. The internal validity of the model may be impaired due to the impact of an uneven distribution of unknown risk modifying characteristics between the

Nomogram	Year of development	Harrell's *C*	95% CI[a]
Stephenson	2006	0.791	0.748–0.833
CAPRA	2005	0.787	0.744–0.831
Kattan	1998	0.744	0.693–0.795

[a]Confidence intervals are calculated using the method described by Newson [22].

TABLE 2
Concordance indices for risk assessment tools when applied to South Australian data

FIG. 2. BCRFP calibration plot (the 45° diagonal line represents a perfect prediction, vertical bars represent 95% confidence intervals).

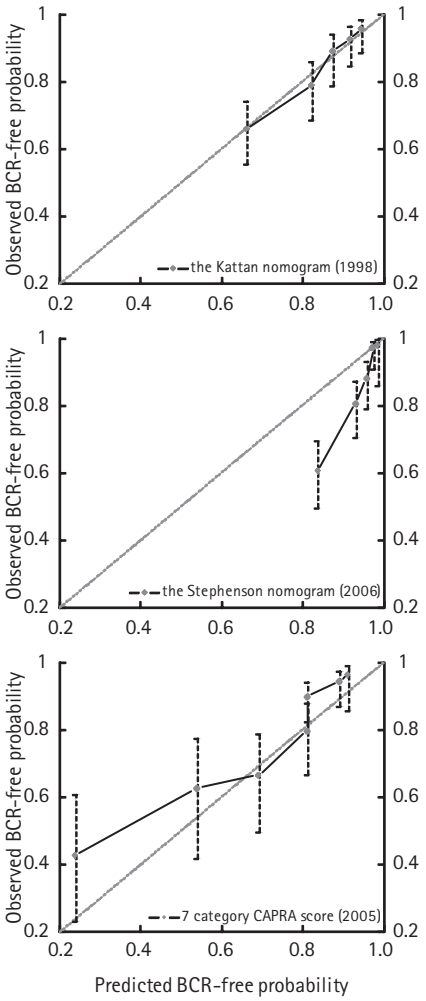
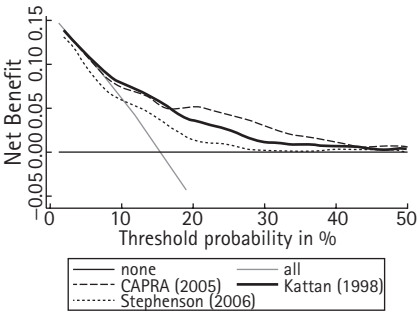


FIG. 3. Decision curve analysis for the prediction of 3-year BCR-free survival. The straight grey line represents the assumption that all patients will recur and the horizontal black line represents the assumption that no patients will recur.



derivation and validation cohorts (confounding) or if variables such as Gleason grade or BCR are measured differently across the populations [18].

There were some differences between the cohorts that were not captured by model variables. Within the PCCOD cohort, BCR was defined as PSA > 0.2 ng/mL or PSA = 0.2 ng/mL given that a subsequent value exceeds 0.2 ng/mL. In contrast, the two nomograms define recurrence as PSA > 0.4 ng/mL. In the derivation of CAPRA, as in the current data set, a threshold of 0.2 ng/mL was used to define a recurrence. Changing the threshold from 0.2 ng/mL to 0.4 ng/mL is likely to increase BCRFP at 3 years without affecting model concordance. However, it is difficult to reconcile the large differences in actuarial and predicted values of BCRFP when altering the definition delays recurrence by a median of only 5 months [24].

In recent years, differences in Gleason score assignment [4] have resulted in an increasing Gleason grade over time [25] and improved Gleason score prediction of prostate cancer outcomes [26]. If changes to pathological grading practices occurred at different times in the USA and Australia, the attributed Gleason score of PCCOD patients may be lower than that of the Stephenson data set despite having identical disease. The risk of PCCOD patients would then be underestimated by the Stephenson nomogram. The Kattan nomogram pre-dated much of the shift in Gleason grade assignment and therefore grading practices may be similar to those used in the database.

This possibility is supported by findings in the PCCOD data set. The mean Gleason score increased from 5.7 in 1998–2003 to 6.5 from 2004 to the present. The ability of Gleason score to predict 3-year BCRFP also improved over this period (receiver–operating characteristic area 0.682 and 0.710 respectively). For change in Gleason grading to explain the disparities between the calibration of the Stephenson and the Kattan nomograms in the PCCOD population, it must have occurred relatively recently in South Australia, perhaps as long as 10 years after it occurred in the USA.

Differences in surgical experience between the model derivation cohorts and the PCCOD cohort may also affect predictive accuracy. This possibility is compelling given that such differences will tend to result in poor calibration without altering concordance, the finding of this current study. Some evidence supports this hypothesis. The Stephenson nomogram was derived from two high volume surgeons at two prestigious US institutions. It is probable that these two surgeons have improved outcomes compared with surgeons who perform fewer prostatectomies [27–29]. The effect of this would be an overestimation of BCRFP when the Stephenson nomogram is applied to other populations, as has been reported in this study. The CAPRA score was derived from a community-based database involving 40 urology practices and, presumably, surgeons of differing experience, similar to the PCCOD.

Our study does have limitations, however, and for these reasons our findings should be considered preliminary. Due to a rapid recent expansion in recruitment, our study is limited by a short median follow-up and it is possible that longer follow-up could influence the findings. However, our observation of marked differences between predicted and actual BCRFP at such an early stage following treatment remains a concerning finding as these differences may well increase.

It is possible that exclusion of patients with missing clinical variables (notably clinical stage) may have introduced a bias. The trend that patients missing data having been treated more recently is a reflection of the recent rapid expansion described above. Patients are often identified following treatment and there is a delay in collecting baseline data. There is no systematic delay for patients with differing characteristics and therefore the exclusion of these patients is unlikely to affect the representativeness of patients included in the analysis.

Further, comparison of patients with clinical stage (included in the analysis) and without clinical stage (excluded from the analysis) showed no differences in age at biopsy, biopsy Gleason, PSA or BCRFP. These are the main predictor variables in the tools being compared and this suggests that the comparison is robust.

CONCLUSIONS

The PCCOD represents a realistic reflection of South Australian, and possibly Australian, RPs. It is therefore reassuring that the concordance of internationally derived risk prediction tools remains high in this Australian cohort. However, if poorly calibrated, the high concordance of the tools is likely to be of little comfort to a physician or the patient being counselled. Our findings support the need for the validation of tools developed in different populations and, if necessary, their recalibration to adjust for the differences in uncaptured risk modifying characteristics of the local population.

ACKNOWLEDGEMENT

We thank the members of the South Australian Prostate Cancer Clinical Outcomes Collaborative for their support of the study.

CONFLICT OF INTEREST

None declared.

REFERENCES

1 **Australian Institute of Health and Welfare and Australasian Association of Cancer Registries.** *Cancer in Australia: an Overview, 2010*, Cancer Series no. 60, Cat. no. CAN 56. Canberra: Australian Institute of Health and Welfare, 2010
2 **Cooperberg MR.** Prostate cancer risk assessment: choosing the sharpest tool in the shed. *Cancer* 2008; **113**: 3062–6
3 **Kattan MW, Eastham JA, Wheeler TM *et al*.** Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors. *J Urol* 2003; **170**: 1792–7
4 **Ohori M, Kattan MW, Koh H *et al*.** Predicting the presence and side of extracapsular extension: a nomogram for staging prostate cancer. *J Urol* 2004; **171**: 1844–9
5 **Stapleton AM, Pinnock CB.** Nomograms for prostate cancer – is their use evidence-based? *Nat Clin Pract Urol* 2005; **2**: 462–3
6 **Cooperberg MR, Pasta DJ, Elkin EP**

***et al*.** The University of California, San Francisco, Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *J Urol* 2005; **173**: 1938–42
7 **Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT.** A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998; **90**: 766–71
8 **Stephenson AJ, Scardino PT, Eastham JA *et al*.** Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Natl Cancer Inst* 2006; **98**: 715–7
9 **Cooperberg MR, Freedland SJ, Pasta DJ *et al*.** Multiinstitutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. *Cancer* 2006; **107**: 2384–91
10 **Graefen M, Karakiewicz PI, Cagiannos I *et al*.** International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2002; **20**: 3206–12
11 **Lughezzani G, Budaus L, Isbarn H *et al*.** Head-to-head comparison of the three most commonly used preoperative models for prediction of biochemical recurrence after radical prostatectomy. *Eur Urol* 2010; **57**: 562–8
12 **May M, Knoll N, Siegsmund M *et al*.** Validity of the CAPRA score to predict biochemical recurrence-free survival after radical prostatectomy. Results from a European multicenter survey of 1,296 patients. *J Urol* 2007; **178**: 1957–62
13 **Zhao KH, Hernandez DJ, Han M, Humphreys EB, Mangold LA, Partin AW.** External validation of University of California, San Francisco, Cancer of the Prostate Risk Assessment score. *Urology* 2008; **72**: 396–400
14 **Eskicorapci SY, Turkeri L, Karabulut E *et al*.** Validation of two preoperative Kattan nomograms predicting recurrence after radical prostatectomy for localized prostate cancer in Turkey: a multicenter study of the Uro-oncology Society. *Urology* 2009; **74**: 1289–95
15 **Graefen M, Karakiewicz PI, Cagiannos I *et al*.** A validation of two preoperative nomograms predicting recurrence following radical prostatectomy in a cohort of European men. *Urol Oncol* 2002; **7**: 141–6
16 **Greene KL, Meng MV, Elkin EP *et al*.** Validation of the Kattan preoperative nomogram for prostate cancer recurrence using a community based cohort: results from cancer of the prostate strategic urological research endeavor (CAPSURE). *J Urol* 2004; **171**: 2255–9
17 **Loeb S, Carvalhal GF, Kan D, Desai A, Catalona WJ.** External validation of the cancer of the prostate risk assessment (CAPRA) score in a single-surgeon radical prostatectomy series. *Urol Oncol* 2010; doi:10.1016/j.urolonc.2010.06.007 (Epub ahead of print)
18 **Kattan MW.** Factors affecting the accuracy of prediction models limit the comparison of rival prediction models when applied to separate data sets. *Eur Urol* 2010; **59**: 566–7
19 **Justice AC, Covinsky KE, Berlin JA.** Assessing the generalizability of prognostic information. *Ann Intern Med* 1999; **130**: 515–24
20 **Schroder F, Kattan MW.** The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. *Eur Urol* 2008; **54**: 274–90
21 **Beckmann K, Pinnock CB, Tamblyn DJ, Kopsaftis T, Stapleton AM, Roder DM.** Clinical and socio-demographic profile of an Australian multi-institutional prostate cancer cohort. *Asia-Pacific J Clin Oncol* 2009; **5**: 247–56
22 **Newson RB.** Comparing the predictive powers of survival models using Harrell's *C* or Somers' *D*. *Stata J* 2010; **10**: 339–58
23 **Vickers AJ, Elkin EB.** Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**: 565–74
24 **Stephenson AJ, Kattan MW, Eastham JA *et al*.** Defining biochemical recurrence of prostate cancer after radical prostatectomy: a proposal for a standardized definition. *J Clin Oncol* 2006; **24**: 3973–8
25 **Freedland SJ, Presti JC Jr, Amling CL *et al*.** Time trends in biochemical recurrence after radical prostatectomy: results of the SEARCH database. *Urology* 2003; **61**: 736–41
26 **Mitchell RE, Shah JB, Desai M *et al*.** Changes in prognostic significance and predictive accuracy of Gleason grading system throughout PSA era: impact of grade migration in prostate cancer. *Urology* 2007; **70**: 706–10
27 **Klein EA, Bianco FJ, Serio AM *et al*.** Surgeon experience is strongly associated with biochemical recurrence after radical prostatectomy for all preoperative risk categories. *J Urol* 2008; **179**: 2212–6
28 **Vickers AJ, Bianco FJ, Serio AM *et al*.** The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst* 2007; **99**: 1171–7
29 **Vickers AJ, Bianco FJ, Gonen M *et al*.** Effects of pathologic stage on the learning curve for radical prostatectomy: evidence that recurrence in organ-confined cancer is largely related to inadequate surgical technique. *Eur Urol* 2008; **53**: 960–6

Correspondence: Carole Pinnock, Repatriation General Hospital, Daws Park, Daw Road, SA 5041, Australia.
e-mail: carole.pinnock@health.sa.gov.au

Abbreviations: **BCRFP**, biochemical-recurrence-free probability; **RP**, radical prostatectomy; **PCCOD**, South Australian Prostate Cancer Clinical Outcomes Database; **BCR**, biochemical recurrence.

EDITORIAL COMMENT

COMPARATIVE ANALYSIS OF THREE RISK ASSESSMENT TOOLS IN AUSTRALIAN PATIENTS WITH PROSTATE CANCER

Tamblyn and colleagues [1] use a population of 635 South Australian radical prostatectomy patients to validate and compare the performance of three published statistical tools [2–4] designed to predict the probability of 3-year post-radical prostatectomy biochemical recurrence (BCR) using preoperative clinical and biopsy data. The authors find that all three statistical tools achieve a similar accuracy in discriminating between South Australian patients in terms of higher or lower risk, producing high concordance indices of 74.4% to 79.1%. However, the performance of the tools varied considerably when calibration was assessed. While the 3-year