available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Prostate Cancer

Predicting Biopsy Outcomes During Active Surveillance for Prostate Cancer: External Validation of the Canary Prostate Active Surveillance Study Risk Calculators in Five Large Active Surveillance Cohorts

Frank-Jan H. Drost [a,b], *Daan Nieboer* [b], *Todd M. Morgan* [c], *Peter R. Carroll* [d], *Monique J. Roobol* [b], *the Movember Foundations Global Action Plan Prostate Cancer Active Surveillance (GAP) Consortium* [1]

[a] *Department of Radiology and Nuclear medicine, Erasmus University Medical Center, Rotterdam, The Netherlands;* [b] *Department of Urology, Erasmus University Medical Center, Rotterdam, The Netherlands;* [c] *Department of Urology, University of Michigan, Ann Arbor, MI, USA;* [d] *Department of Urology, Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA*

Article info

Article history:
Accepted July 24, 2019

Associate Editor:
Matthew Cooperberg

Statistical Editor:
Andrew Vickers

Keywords:
Active surveillance
Decision curve analyses
Disease reclassification
External validation
Low-risk prostate cancer
Net benefit
Prostate biopsy
Risk calculator

Abstract

Background: Men with prostate cancer (PCa) on active surveillance (AS) are followed through regular prostate biopsies, a burdensome and often unnecessary intervention, not without risks. Identifying men with at a low risk of disease reclassification may help reduce the number of biopsies.
Objective: To assess the external validity of two Canary Prostate Active Surveillance Study Risk Calculators (PASS-RCs), which estimate the probability of reclassification (Gleason grade $\geq$7 with or without >34% of biopsy cores positive for PCa) on a surveillance biopsy, using a mix of months since last biopsy, age, body mass index, prostate-specific antigen, prostate volume, number of prior negative biopsies, and percentage (or ratio) of positive cores on last biopsy.
Design, setting, and participants: We used data up to November 2017 from the Movember Foundation's Global Action Plan (GAP3) consortium, a global collaboration between AS studies.
Outcome measurements and statistical analysis: External validity of the PASS-RCs for estimating reclassification on biopsy was assessed by calibration, discrimination, and decision curve analyses.
Results and limitations: Five validation cohorts (Prostate Cancer Research International: Active Surveillance, Johns Hopkins, Toronto, Memorial Sloan Kettering Cancer Center, and University of California San Francisco), comprising 5105 men on AS, were eligible for analysis. The individual cohorts comprised 429–2416 men, with a median follow-up between 36 and 84 mo, in both community and academic practices mainly from western countries. Abilities of the PASS-RCs to discriminate between men with and without reclassification on biopsy were reasonably good (area under the receiver operating characteristic curve values 0.68 and 0.65). The PASS-RCs were moderately well calibrated, and had a greater *net* benefit than most default strategies between a predicted 10% and 30% risk of reclassification.

[1] See Appendix for Members of The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium.

* Corresponding author at: Office NA-1520, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.
E-mail address: f.drost@erasmusmc.nl (F.H. Drost).

https://doi.org/10.1016/j.eururo.2019.07.041

Conclusions: Both PASS-RCs improved the balance between detecting reclassification and performing surveillance biopsies by reducing unnecessary biopsies. Recalibration to the local setting will increase their clinical usefulness and is therefore required before implementation.
Patient summary: Unnecessary prostate biopsies while on active surveillance (AS) should be avoided as much as possible. The ability of two calculators to selectively identify men at risk of progression was tested in a large cohort of men with low-risk prostate cancer on AS. The calculators were able to prevent unnecessary biopsies in some men. Usefulness of the calculators can be increased by adjusting them to the characteristics of the population of the clinic in which the calculators will be used.

1. Introduction

As many as two-thirds of men with newly diagnosed prostate cancer are considered to have a low risk of progression [1]. For these men, active surveillance (AS) is increasingly recognised as a favourable alternative to immediate, radical therapy [2]. Although AS can safely reduce the harmful overtreatment of low-risk disease, it still entails relatively frequent surveillance prostate biopsies that are costly, potentially unnecessary, and harmful [3].

There is wide variability in how AS has been implemented. Surveillance biopsies can be scheduled annually, every 2–4 yr, or can be triggered after an indirect sign of disease progression (eg, by prostate-specific antigen [PSA] kinetics) [4]. However, the outcomes of up to 64–90% of biopsy procedures do not lead to a change in treatment strategy [3,5]. In addition, every prostate biopsy is considered bothersome and is associated with a 0.5–6.9% risk of serious complications [6]. In clinical practice, this may lead to a considerable decrease in compliance for scheduled biopsies [7,8], especially after a previous biopsy complication [9]. Remarkably, after an indirect sign of disease progression based on PSA kinetics, when a surveillance biopsy is potentially most important, the biopsy rate is <30% [7]. These observations indicate the need to develop more accurate and personalised monitoring methods that safely result in fewer unnecessary biopsies, rather than the "one size fits all" or the "single" trigger-dependent strategies that are currently used.

To aid in personalised monitoring, the Canary Prostate Active Surveillance Study Risk Calculator (PASS-RC) [10] was developed in 2015 and recently updated [11]. The PASS-RCs estimate the risk of two definitions of disease progression at surveillance biopsy, based on different sets of interacting and readily available parameters. They are made available online (at the Canary PASS website) as tools that can be used to guide decision-making regarding the performance of a surveillance biopsy.

Before any risk calculator is broadly implemented in clinical practice, external validation is needed. Since 2016, many of the world's largest AS cohort databases have been integrated through the Movember Foundation's Global Action Plan (GAP3) AS project [12]. The objective of this study was to externally validate the PASS-RCs using data from this GAP3 consortium. We focused on the original PASS-RC because the required set of parameters was available in all our validation cohorts and reported the results of the new PASS-RC in a subset of cohorts in the Supplementary material.

2. Patients and methods

2.0.1. Study population

We selected five AS cohorts from GAP3 based on the number of events (>100 reclassifications on biopsy) and applicability to ensure reliable validation results. Similar to the development cohort [10], the patient inclusion criteria were as follows: Gleason ≤6 localised cT1–2 prostate cancer, initial PSA < 20 ng/ml, <34% of the biopsy cores positive for cancer, and one or more postdiagnostic biopsies. Patients from the PASS cohorts who are also included in GAP3 were excluded. Biopsy results were excluded when data were incomplete, when no PSA within 12 mo prior to the biopsy procedure was measured, and when PSA appeared to be spuriously high (>20 ng/ml with lower values before and after). All other biopsy results were taken into account, irrespective of the biopsy method (ie, systematic biopsy or possibly taken magnetic resonance imaging [MRI]-guided biopsy) as the difference was not recorded. Table 1 shows the variety of follow-up protocols of the development cohorts and the five validation cohorts [5,13–21]. All cohorts and all biopsy procedures attributed to the assessment of the performance of the original PASS-RC. The Toronto cohort, however, could not be used in the validation of the new PASS-RC because data on prostate volume were not available. Furthermore, the performance of the new PASS-RC could be assessed only for confirmatory biopsy procedures because data on prostate volume at subsequent surveillance biopsies were not available.

2.0.2. Risk factors and outcomes

The PASS-RCs are logistic regression models. The original PASS-RC is based on age at biopsy, months since last biopsy, last PSA value, one or more prior negative biopsies, and the percentage of positive cores on the last biopsy [10]. Its target condition, disease reclassification, was defined as a Gleason score upgrade from ≤6 to ≥7 or an increase in percentage of biopsy cores positive for cancer from <34% to ≥34%. The new PASS-RC is based on age at biopsy, body mass index

Table 1 – Follow-up protocols of the cohorts

	Cohort[a]	PSA (mo)	Confirmatory biopsy (mo)	Repeat biopsies (yr from previous)	Triggers for biopsy
Development cohort	Canary-PASS	3	0–12	2	–
Validation cohorts	Johns Hopkins	6	<12	1	–
	MSKCC	6	3	First 1–1.5, then 2–3	DRE change or sustained PSA increase
	PRIAS	3 (for 2 yr), then 6	≤12	3	PSA-DT 3–10 yr
	Toronto	3 (for 2 yr), then 6	≤12	3–4	PSA-DT <3 yr[b]
	UCSF	3	<12	1–2	–

Canary-PASS = Canary Prostate Active Surveillance Study; DRE = digital rectal examination; MSKCC = Memorial Sloan Kettering Cancer Center; PRIAS = Prostate cancer Research International: Active Surveillance; PSA = prostate-specific antigen; PSA-DT = PSA doubling time; UCSF = University of California San Francisco.
[a] The role of magnetic resonance imaging (and biomarkers) during the study period in these cohorts was not specified.
[b] As of 2009.

(BMI), positive core ratio >0.2, two or more prior negative biopsies, log (prostate volume), and log (PSA) [11]. We used the published odds ratios and the new PASS-RC to derive the model intercept (Canary PASS website accessed: 25 January 2019). Its target condition was defined as a Gleason score upgrade from ≤6 to ≥7, without any cancer volume criteria.

2.0.3. *Calibration and discrimination*

Calibration and discrimination of the PASS-RCs were assessed for each validation cohort separately. Calibration addresses the agreement between the predicted probabilities and observed outcomes, that is, the reliability of the estimated risks. Calibration was quantified by the calibration in the large (calibration-i.t.l.) and the calibration slope, and assessed graphically [22]. The calibration-i.t.l. indicates whether the predicted probabilities are systematically too low or too high, and should ideally be zero. The calibration slope indicates whether the average effects of the predictors used in the PASS-RC are on average too weak or too strong, and should ideally be 1. The discriminative ability, between men with and without reclassification on biopsy, was quantified by calculating the area under the receiver operating characteristic curve (AUC). The degree of variety in patient characteristics (case-mix

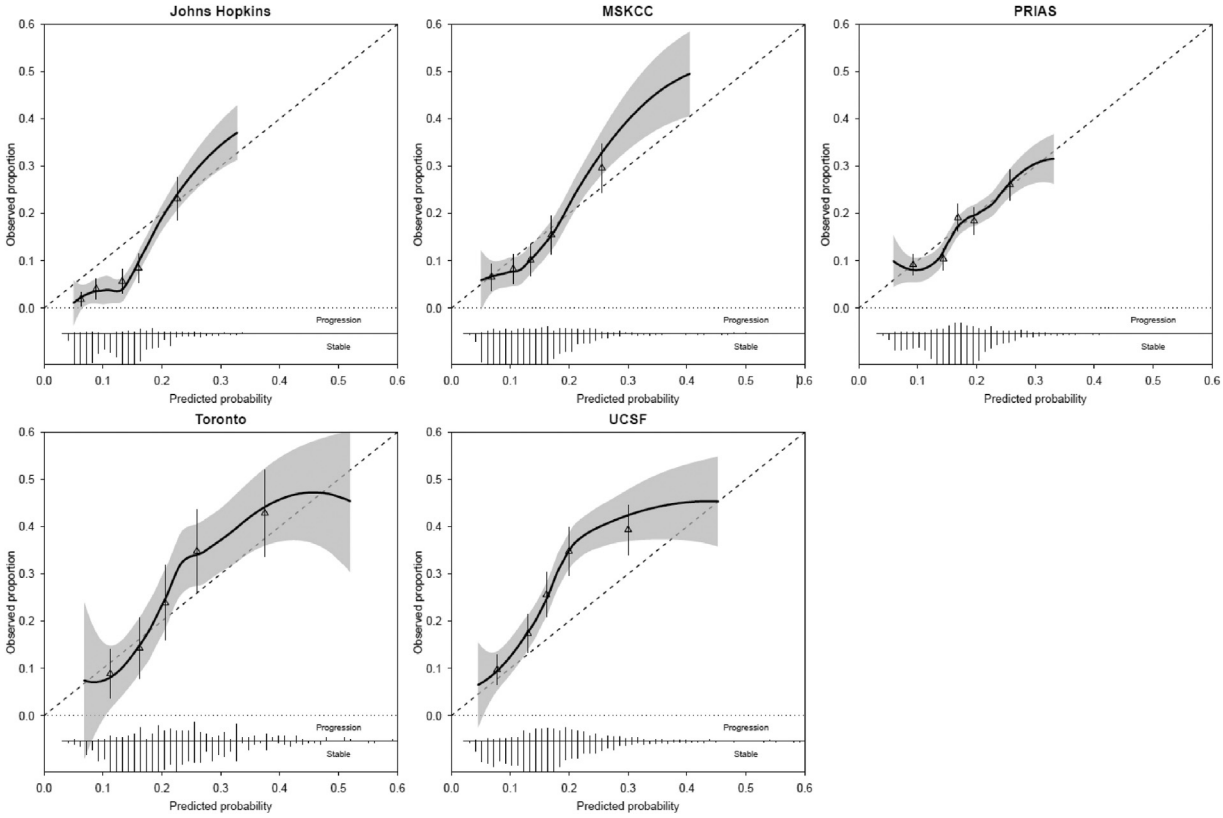

Fig. 1 – Calibration plots of the original PASS-RC[a]. MSKCC = Memorial Sloan Kettering Cancer Center; PASS-RC = Prostate Active Surveillance Study Risk Calculator; PRIAS = Prostate Cancer Research International: Active Surveillance; UCSF = University of California San Francisco.[a] Calibration plots depicting the agreement between the predicted probabilities calculated by the original PASS-RC and the observed outcomes in the five validation cohorts.

heterogeneity) influences the ability to distinguish a subgroup on the basis of these characteristics. Case-mix heterogeneity was therefore quantified using the standard deviation of the linear predictor of the prediction models [23].

2.0.4. Clinical usefulness

Clinical usefulness was assessed using decision curve analysis [24]. We assessed the ability of the PASS-RCs to make better selection of men who need a biopsy compared with the default strategies of the validation cohorts as described in Table 1. "Net" benefits were estimated by summing the benefits (detecting reclassification) and subtracting the harms (performing unnecessary biopsies), weighted by a factor related to how many men a physician is willing to biopsy in order to find one reclassification. For example, if a physician (in a shared decision with the patient) is willing to biopsy 10 men to detect one man with reclassification, it implies that the harm of missing a reclassification is considered nine times greater than that of performing an unnecessary biopsy (in 10 men undergoing biopsy, one reclassification detected equates to nine unnecessary biopsies for each detected reclassification). In this case, the weighing factor would be 0.11 (1/9), indicating the so-called "risk threshold" at or above which a biopsy would be performed. However, as such a risk threshold is subjective to the preferences of physicians and patients, a range of risk thresholds needs to be investigated. The *net* benefit was plotted for each validation cohort against a range of clinically relevant risk thresholds. At the risk thresholds where the PASS-RCs have a higher *net* benefit than the default strategy and the strategy of not performing any biopsies, the PASS-RCs are considered clinically useful. In addition, we quantified the *net* benefit by calculating the "net" number of reduced biopsies when using the PASS-RCs. The *net* number of reduced biopsies considers that a negative biopsy predicted by the PASS-RCs can be a true or a false negative, and adjusts for the false negatives. Therefore, the *net* number of reduced biopsies is equivalent to the number of biopsies that would be reduced without missing a reclassification.

All statistical analyses were performed using R version 3.3.0 (R Foundation for Statistical Computing, Vienna, Austria) using data from the start of the cohort studies up to November 2017. Missing values for BMI and prostate volume were imputed using the *mice* package.

3. Results

3.0.1. Study population

The five validation cohorts comprised a total of 5105 men eligible for analyses, with considerable differences in characteristics (Table 2). Twelve, 48, and 162 men were excluded because of a spuriously high PSA value, because of no PSA within 1 yr of biopsy, and because they were also included in the PASS development cohorts, respectively. Compared with the development cohorts, the age at diagnosis in the Toronto, Johns Hopkins, and Prostate Cancer Research International: Active Surveillance (PRIAS) patients was higher (62 vs 65, 66, and 66 yr, respectively). PSA values of the patients from the Johns Hopkins and Memorial Sloan Kettering Cancer Center (MSKCC) cohorts were lower (5.5 vs 4.8 and 4.6 ng/ml, respectively). There

Table 2 – Cohort characteristics

Cohorts	Development cohorts		Validation cohorts				
	Original PASS-RC	New PASS-RC	Toronto	Johns Hopkins	PRIAS[a]	UCSF	MSKCC
No. of patients	859	558	429	651	2416	864	745
Age at diagnosis (yr), median (IQR)	62	NS	65 (60–70)	66 (62–69)	66 (61–71)	62 (57–67)	63 (57–68)
Body mass index (kg/m^2), median (IQR)	NS	NS	NS	26 (25–29)	26 (24–28)	27 (25–30)	27 (25–31)
PSA[b] (ng/ml), median (IQR)	5.5 (3.0)[c]	NS	5.4 (3.8–7.2)	4.8 (3.6–6.2)	5.5 (4.1–7.3)	5.4 (4.2–7.3)	4.6 (3.4–6.2)
Prostate volume (cm^3), median (IQR)	NS	41 (28–54)	NS	45 (36–60)	38 (28–52)	45 (35–59)	48 (37–63)
Follow-up (mo), median (IQR)	36 (3–149)[d]	NS	84 (54–105)	48 (26–75)	36 (22–54)	50 (30–85)	62 (47–82
No. of surveillance biopsies per patient (%)							
Confirmatory biopsy	859 (100)	319 (57)	429 (100)	651 (100)	2416 (100)	864 (100)	745 (100)
One subsequent biopsy	458 (53)	246 (44)	117 (27)	368 (57)	759 (31)	423 (49)	459 (62)
Two or more subsequent biopsies	211 (25)	108 (19)	15 (3)	244 (37)	236 (10)	204 (24)	204 (27)
CR for diagnostic biopsy[b], median (IQR)	NS	0.08 (0.08)	NS	0.08 (0.08)	0.10 (0.08)	0.10 (0.10)	0.08 (0.08)
Total reclassified patients (GS ≥7 or ≥34% positive cores)	41%	41%[e]	33%	21%	24%	48%	29%
Mean reclassified patients per biopsy round	21%	18%	25%	9%	17%	25%	14%
Owing to GS ≥7	75%	100%	74%	82%	73%	78%	87%
Owing to ≥34% positive cores	36%	0%	59%	35%	48%	44%	36%

CR = core ratio; GS = Gleason score; IQR = interquartile range; MSKCC = Memorial Sloan Kettering Cancer Center; NS = not specified; PASS-RC = Prostate Active Surveillance Study Risk Calculator; PRIAS = Prostate Cancer Research International: Active Surveillance; PSA = prostate-specific antigen; UCSF = University of California San Francisco.

[a] PRIAS comprised 148 academic or community practices from 18 countries.

[b] CR is defined as the number of biopsy cores containing cancer divided by the total number of biopsy cores in the diagnostic biopsy.

[c] Mean (SD).

[d] Median (range).

[e] Of the 319 patients who underwent a confirmatory biopsy.

were less reclassified patients in the Johns Hopkins and PRIAS cohorts (41% vs 21% and 24%, respectively), and the duration of follow-up differed substantially in almost all cohorts. Johns Hopkins and MSKCC cohorts had lower mean reclassification rates per biopsy round (9% and 14%, respectively) than both development cohorts (21% and 18%). Detailed cohort characteristics per biopsy round are presented in Supplementary material (mmc2).

3.0.2. Calibration and discrimination

The original PASS-RC showed reasonable calibration in most cohorts Fig. 1. However, the risk of reclassification on biopsy was overestimated in the Johns Hopkins cohort and underestimated in the University of California San Francisco (UCSF) cohort (calibration-i.t.l. –0.51 and 0.51, respectively; Fig. 2). The effects of the predictors were too weak in the Johns Hopkins cohort (calibration slope 1.72). The average AUC of the PASS-RC was 0.68 (95% confidence interval [CI] 0.64–0.72), varying between 0.63 and 0.75 (Fig. 2).

The new PASS-RC overestimated the risk of reclassification at confirmatory biopsy in most cohorts (pooled calibration-i.t.l. –0.27 [95% CI –0.57 to 0.03]; Supplementary Fig. 1 and 2). Calibration slopes of all cohorts were well below 1, indicating too strong predictor effects. The average AUC of the new PASS-RC was 0.65 (95% CI 0.61–0.68), varying between 0.62 and 0.69.

Supplementary Fig. 3 (mmcX) shows that the degree of variety in patient characteristics may partially have influenced the discriminative ability of the PASS-RCs.

3.0.3. Clinical usefulness

The decision curves (Fig. 3) show a higher *net* benefit in all cohorts when using the original PASS-RC for a varying risk-threshold range of reclassification between 10% and 30%, compared with the default strategy for selecting men who need a biopsy ("all") and the strategy of not performing any biopsies ("none"). Despite a varying *net* benefit within a varying risk-threshold range, it seems that the original PASS-RC is clinically useful and its use would lead to more efficient detection of reclassification in most cohorts.

In the assessment of the new PASS-RC (mmcX in the Supplementary materials), the decision curves show that the new PASS-RC might also be clinically useful in some cohorts. However, the *net* benefit and the risk-threshold range in which the *net* benefit was higher seemed limited. Despite that, the use of the new PASS-RC might lead to more efficient detection of reclassification in some cohorts.

The number of biopsies that can be avoided without missing an extra reclassification by using the PASS-RCs instead of the default strategies is presented per risk threshold (for the original PASS-RC in Table 3 and for the new PASS-RC in Supplementary Table 2 (mmcX)). With both PASS-RCs, the reduction of avoidable biopsies was highest in the Johns Hopkins cohort and lowest in the UCSF cohort. It should be noted that these numbers are the *net* numbers, calculated from the absolute numbers of missed reclassifications and saved biopsies, while accounting for how many men a physician is willing to biopsy in order to find one reclassification (the risk threshold). Hence, although the use of the PASS-RCs implies missing a few reclassifications, the balance between missed reclassifications and unnecessary biopsies seems more favourable as compared with the default strategy at any risk threshold between 10% and 30% in most cohorts. For example, at a risk threshold of 15%, the numbers of *net* reduced biopsies per 100 biopsies ranged between cohorts from 3 to 47 with the original PASS-RC and from 12 to 33 with the new PASS-RC.

4. Discussion

In order to reduce potentially unnecessary, costly, and harmful prostate biopsies during AS, the PASS-RCs were developed to guide prostate biopsy decision-making. In this study, we assessed the usefulness and external validity of the PASS-RCs for predicting reclassification on biopsy in five large AS cohorts from the Movember Foundation's GAP3 consortium.

The original PASS-RC was validated in all available biopsies and in all cohorts. The new PASS-RC used different

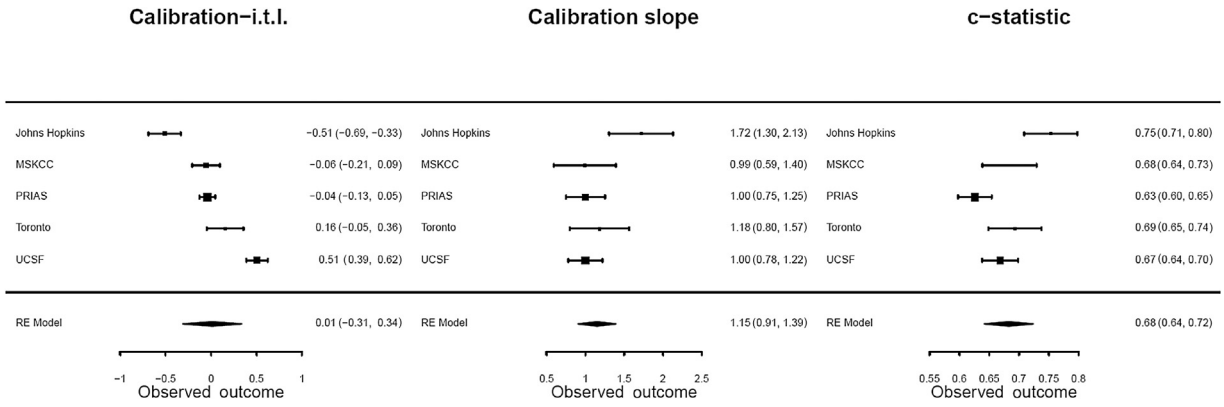

Fig. 2 – Calibration metrics and area under the curve of the original PASS-RC[a]. Calibration-i.t.l. = calibration in the large; c-statistic = area under the curve; MSKCC = Memorial Sloan Kettering Cancer Center; PASS-RC = Prostate Active Surveillance Study Risk Calculator; PRIAS = Prostate cancer Research International: Active Surveillance; RE = random effect; UCSF = University of California San Francisco.[a] Diamonds and brackets indicate the pooled summary estimate 95% confidence intervals; horizontal lines and brackets indicate individual cohort 95% confidence intervals.

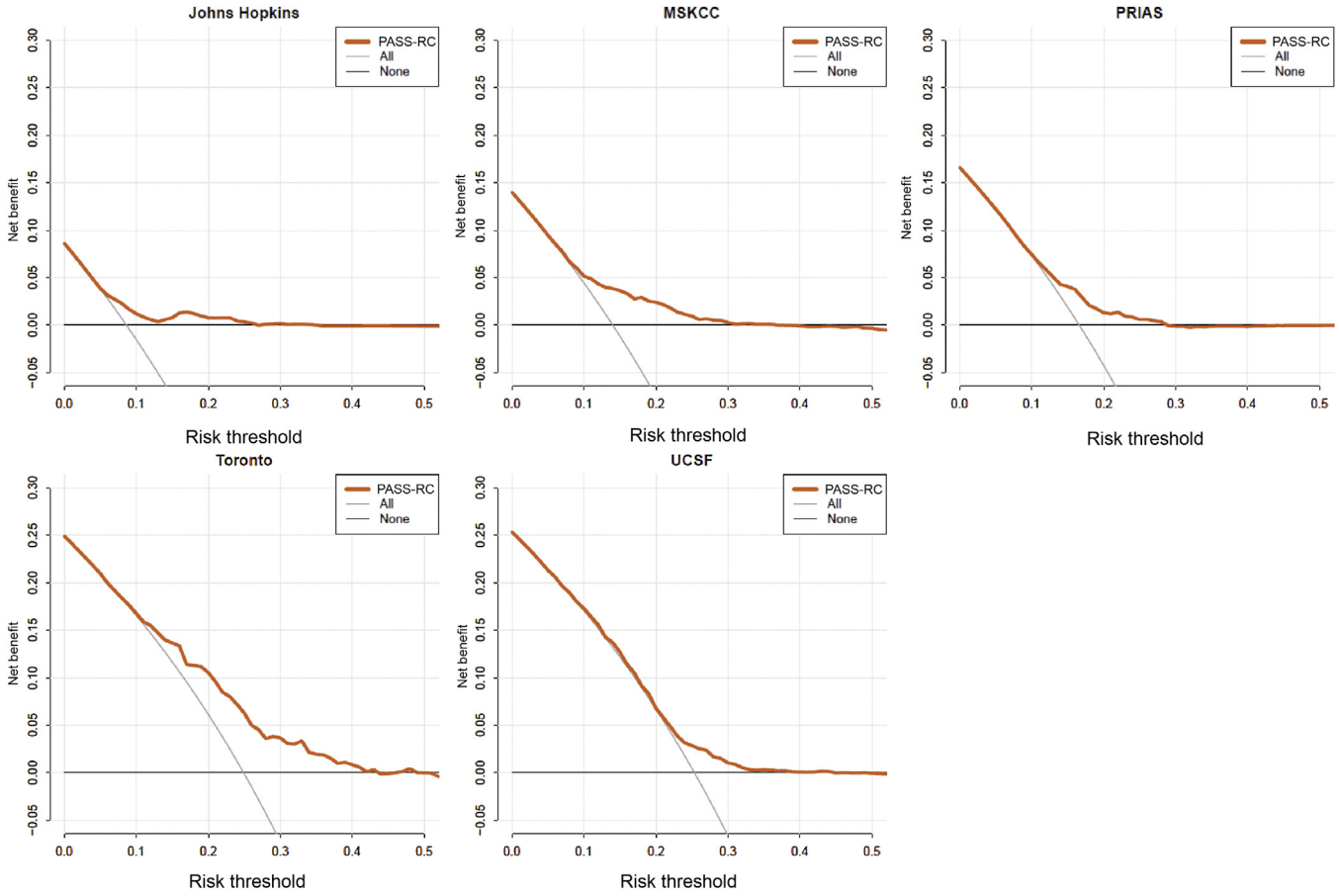

Fig. 3 – Decision curve analyses of the original PASS-RC[a]. MSKCC = Memorial Sloan Kettering Cancer Center; PASS-RC = Prostate Active Surveillance Study Risk Calculator; PRIAS = Prostate Cancer Research International: Active Surveillance; UCSF = University of California San Francisco. [a] Decision curve analyses showing the *net* benefit of the original PASS-RC in reference to the default biopsy strategies ("all") and not biopsying any man ("none") for detecting reclassification across all threshold probabilities in the six validation cohorts. Choosing a risk threshold of 0.15 indicates that the physician and the patient are willing to biopsy 7.67 times (1 + [1/0.15]) in order to find one reclassification.

parameters, which were not always available in the cohorts, and could therefore be validated only in confirmatory biopsies and not in the Toronto cohort. Regardless of these differences, both PASS-RCs seemed clinically useful.

The PASS-RCs estimated probabilities of reclassification on biopsy, which were in reasonable agreement with the true outcomes in most of the cohorts. However, there were substantial over- and underestimations of the true outcomes in some cohorts. The differences in patient selection, follow-up protocols, and resulting reclassification rates between the development and validation cohorts might explain some of the variable performance of the PASS-RCs. The risk of reclassification was overestimated in the Johns Hopkins cohort, possibly because of stringent low-risk inclusion criteria, frequent performance of biopsies (yearly vs every other year), and subsequent low reclassification

Table 3 – Reduction of avoidable biopsies per 100 biopsies by using the original PASS-RC

Risk thresholds (%)	Johns Hopkins	MSKCC	PRIAS	Toronto	UCSF
5	1	0	0	0	–2
10	24	6	1	2	2
15	47	27	12	11	3
20	60	40	22	18	0
25	67	47	35	19	7
30	72	54	44	26	18

MSKCC = Memorial Sloan Kettering Cancer Center; PASS-RC = Prostate Active Surveillance Study Risk Calculator; PRIAS = Prostate Cancer Research International: Active Surveillance; UCSF = University of California San Francisco.
Clinical utility is quantified by calculating the *net* number of biopsies that can be avoided by using the original PASS-RC, as compared with the default strategies of the validation cohorts. This number is given per risk threshold for reclassification and is equivalent to the reduction in unnecessary biopsies without missing an extra reclassification.

rates per biopsy round (9% vs 21%). In contrast, the relatively high reclassification rates per biopsy round in the UCSF cohort (25%) might explain the underestimated risk of reclassification in this cohort. However, it is likely that interacting factors and unregistered factors (eg, the use of MRI and other biomarkers) contributed to the variable performance of the PASS-RCs between cohorts. In general, all substantial differences regarding the selection and follow-up of men between the development cohort and the cohort in which a PASS-RC is to be used can result in miscalibrations and a reduced *net* benefit. Therefore, recalibrating the PASS-RCs to individual cohorts is needed to avoid over- or underestimation of the risk of reclassification, which otherwise would result in performing biopsies in too many or too few men, respectively.

The ability of the original PASS-RC to discriminate between men with and without reclassification ranged between AUCs of 0.63 and 0.75. The discriminative ability of the new PASS-RC ranged between AUCs of 0.62 and 0.69, with lower AUC values in both cases as compared with those of the development cohorts. The discriminative ability partially depends on case-mix heterogeneity within cohorts, that is, the variability in patient characteristics. In other words, when men have very similar characteristics, it is difficult to distinguish a subgroup on the basis of these characteristics.

Using the PASS-RCs to select men for biopsy instead of the default strategy would lead to a higher *net* benefit in most cohorts. However, the *net* benefit was higher only for a varying risk-threshold range between 10% and 30% across cohorts. Therefore, the clinical usefulness of the PASS-RCs also depend on the risk tolerance of physicians and patients (ie, some might accept a higher risk of missing [or delaying] a reclassification, in exchange for not performing a biopsy, than others). If patient and physician are willing to accept a risk of 10–30%, then the PASS-RCs outperform most default strategies with regard to detecting reclassification and avoiding unnecessary biopsies. To our knowledge, potential risk thresholds used for decision-making in this setting have not yet been explored.

Quantifying the *net* benefit in terms of the avoidable number of unnecessary biopsies demonstrated a large variability between PASS-RCs and cohorts. At a 15% risk threshold, the original PASS-RC would be equivalent to a strategy that reduces 3–47% of biopsies without missing one extra reclassification compared with the default strategies. The new PASS-RC would reduce 12–33% of biopsies. Reducing the number of unnecessary biopsies could lead to improved adherence to AS protocols [7]. Both PASS-RCs require readily available clinical parameters and are available online; therefore, no extra diagnostic tests are needed and no extra costs are incurred.

The calibration-i.t.l. showed large heterogeneity between the validation cohorts, reflecting differences in baseline risks of reclassification on biopsy. Miscalibration of the PASS-RCs might lead to too many or too few biopsies. Therefore, recalibration by a local statistician is advised before applying the PASS-RCs to a new setting [25]. Recalibration will result in higher percentages of *net* reduced biopsies across a broader set of risk thresholds and would reduce the risk of lower *net* benefits than the default strategies [26]. In other words, recalibration would increase their clinical usefulness and would minimise the risk of an unfavourable balance between missed reclassifications and unnecessary biopsies.

Both PASS-RCs are based on data largely from before 2015, and as the authors acknowledge, new updates with longer follow-up are needed to improve the PASS-RCs [10], possibly incorporating information from MRI and molecular biomarkers [27,28]. Furthermore, the new PASS-RC was developed in a relatively small cohort of 558 men (of whom only 319 had one or more surveillance biopsies) because the remaining men were either used for internal validation or excluded (mainly based on unavailability of kallikrein data that were used for another version of the new PASS-RC). This selection of men might have influenced the performance of the new PASS-RC.

The new PASS-RC used reclassification from GS 6 to $\geq$7 without any volume criteria as the target condition, which might be more clinically useful than the one used in the original PASS-RC (including the criterion of $\geq$34% positive biopsy cores). However, other definitions of reclassification could be justified, as there is little consensus on the best criteria for selecting men for immediate treatment [4]. Owing to missing sequential prostate volume data in the validation cohorts, we were able to validate the new PASS-RC at confirmatory biopsy only. Clinicians should take these differences into account when making a decision to use one of the PASS-RCs.

Another study validated the original PASS-RC with other validation methods than those used in this study [29]. It used a multicentre Spanish cohort and showed results within the variability of the GAP3 cohort results.

4.1. *Strengths and limitations*

The five GAP3 cohorts used for this validation provide a unique reference for the PASS-RCs to explore their usability in a wide range of clinical settings. The differences between patient selection and follow-up protocols comprise almost all possible combinations of present AS strategies worldwide. The external validation method used in this study is widely acknowledged to be the most reliable method to ensure proper assessment of prediction models [24,30,31]. However, the relatively short follow-up of some cohorts in combination with the inevitable inclusion criteria for this study (one or more postdiagnostic biopsies) could have led to a selection bias, especially when considering that surveillances biopsies are frequently performed later than planned according to the protocols used by most centres [32]. Hence, high-risk patients might have been included (and also sampled) more frequently than low-risk patients. Furthermore, we acknowledge that we could not fully explain the variable performance of the PASS-RCs between cohorts, possibly due to interacting and unregistered differences in the selection and follow-up of patients.

5. Conclusions

The PASS-RCs demonstrated good discriminative ability and overall calibration in most cohorts assessed here. These

validation data indicate that PASS-RC may be incorporated into clinical practice as a tool for improving selection of men for surveillance biopsy. The use of this tool could reduce unnecessary biopsies, potentially improving adherence to AS protocols. Since calibration was suboptimal in most cohorts, recalibration of both PASS-RC models prior to implementation in a new setting would substantially improve their clinical usefulness, and would minimise the risk of an unfavourable balance between missed reclassifications and unnecessary biopsies.

Author contributions: Frank-Jan H. Drost had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Drost, Nieboer, Roobol.
Acquisition of data: Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) Consortium.
Analysis and interpretation of data: Drost, Nieboer.
Drafting of the manuscript: Drost.
Critical revision of the manuscript for important intellectual content: Drost, Nieboer, Roobol, Morgan, Carroll.
Statistical analysis: Nieboer.
Obtaining funding: Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) Consortium.
Administrative, technical, or material support: Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) Consortium.
Supervision: Roobol.
Other: None.

Financial disclosures: Frank-Jan H. Drost certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: This work was supported by the Movember Foundation. The funder did not play any role in the study design, collection, analysis, or interpretation of data, or in the drafting of this paper.

Appendix. Members of The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium

Principle Investigators: Bruce Trock (Johns Hopkins University, The James Buchanan Brady Urological Institute, Baltimore, USA), Behfar Ehdaie (Memorial Sloan Kettering Cancer Center, New York, USA), Peter Carroll (University of California San Francisco, San Francisco, USA), Christopher Filson (Emory University School of Medicine, Winship Cancer Institute, Atlanta, USA), Jeri Kim / Christopher Logothetis (MD Anderson Cancer Centre, Houston, USA), Todd Morgan (University of Michigan and Michigan Urological Surgery Improvement Collaborative (MUSIC), Michigan, USA), Laurence Klotz (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada), Tom Pickles (University of British Columbia, BC Cancer Agency, Vancouver, Canada), Eric Hyndman (University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada), Caroline M. Moore (University College London & University College London Hospital Trust, London, UK), Vincent Gnanapragasam (University of Cambridge & Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK), Mieke Van Hemelrijck (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Prokar Dasgupta (Guy's and St Thomas' NHS Foundation Trust, London, UK), Chris Bangma (Erasmus Medical Center, Rotterdam, The Netherlands/ representative of Prostate cancer Research International Active Surveillance (PRIAS) consortium), Monique Roobol (Erasmus Medical Center, Rotterdam, The Netherlands/ representative of Prostate cancer Research International Active Surveillance (PRIAS) consortium), Arnauld Villers (Lille University Hospital Center, Lille, France), Antti Rannikko (Helsinki University and Helsinki University Hospital, Helsinki, Finland), Riccardo Valdagni (Department of Oncology and Hemato-oncology, Università degli Studi di Milano, Radiation Oncology 1 and Prostate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy), Antoinette Perry (University College Dublin, Dublin, Ireland), Jonas Hugosson (Sahlgrenska University Hospital, Göteborg, Sweden), Jose Rubio-Briones (Instituto Valenciano de Oncología, Valencia, Spain), Anders Bjartell (Skåne University Hospital, Malmö, Sweden), Lukas Hefermehl (Kantonsspital Baden, Baden, Switzerland), Lee Lui Shiong (Singapore General Hospital, Singapore, Singapore), Mark Frydenberg (Monash Health; Monash University, Melbourne, Australia), Yoshiyuki Kakehi (Kagawa University Faculty of Medicine, Kagawa, Japan), Byung Ha Chung (Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea)

Pathologist: Theo van der Kwast (Princess Margaret Cancer Centre, Toronto, Canada).

Technology Research Partners: Wim van der Linden (Royal Philips, Eindhoven, the Netherlands), Tim Hulsen (Royal Philips, Eindhoven, the Netherlands), Cees de Jonge (Royal Philips, Eindhoven, the Netherlands).

(Advisory) Regional statisticians: Mike Kattan (Cleveland Clinic, Cleveland, Ohio, USA), Ji Xinge (Cleveland Clinic, Cleveland, Ohio, USA), Kenneth Muir (University of Manchester, Manchester, UK), Artitaya Lophatananon (University of Manchester, Manchester, UK), Michael Fahey (Epworth HealthCare, Melbourne, Australia), Ewout Steyerberg (Erasmus Medical Center, Rotterdam, The Netherlands), Daan Nieboer (Erasmus Medical Center, Rotterdam, The Netherlands); Liying Zhang (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada)

*(executive) Regional statisticians:*Ewout Steyerberg (Erasmus Medical Center, Rotterdam, The Netherlands), Daan Nieboer (Erasmus Medical Center, Rotterdam, The Netherlands); Kerri Beckmann (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Brian Denton (University of Michigan, Michigan, USA), Andrew Hayen (University of Technology

Sydney, Australia), Paul Boutros (Ontario Institute of Cancer Research, Toronto, Ontario, Canada).

Clinical Research Partners' IT Experts: Wei Guo (Johns Hopkins University, The James Buchanan Brady Urological Institute, Baltimore, USA), Nicole Benfante (Memorial Sloan Kettering Cancer Center, New York, USA), Janet Cowan (University of California San Francisco, San Francisco, USA), Dattatraya Patil (Emory University School of Medicine, Winship Cancer Institute, Atlanta, USA), Emily Tolosa (MD Anderson Cancer Centre, Houston, Texas, USA), Tae-Kyung Kim (University of Michigan and Michigan Urological Surgery Improvement Collaborative, Ann Arbor, Michigan, USA), Alexandre Mamedov (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada), Vincent LaPointe (University of British Columbia, BC Cancer Agency, Vancouver, Canada), Trafford Crump (University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada), Jenna Kimberly-Duffell (University of Cambridge & Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK), Aida Santaolalla (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Daan Nieboer (Erasmus Medical Center, Rotterdam, The Netherlands), Jonathan Olivier (Lille University Hospital Center, Lille, France), Tiziana Rancati (Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy), Helén Ahlgren (Sahlgrenska University Hospital, Göteborg, Sweden), Juanma Mascarós (Instituto Valenciano de Oncología, Valencia, Spain), Annica Löfgren (Skåne University Hospital, Malmö, Sweden), Kurt Lehmann (Kantonsspital Baden, Baden, Switzerland), Catherine Han Lin (Monash University and Epworth HealthCare, Melbourne, Australia), Hiromi Hirama (Kagawa University, Kagawa, Japan), Kwang Suk Lee (Yonsei University College of Medicine, Gangnam Severance Hospital, Seoul, Korea).

Research Advisory Committee: Guido Jenster (Erasmus MC, Rotterdam, the Netherlands), Anssi Auvinen (University of Tampere, Tampere, Finland), Anders Bjartell (Skåne University Hospital, Malmö, Sweden), Masoom Haider (University of Toronto, Toronto, Canada), Kees van Bochove (The Hyve B.V. Utrecht, Utrecht, the Netherlands), Ballentine Carter (Johns Hopkins University, Baltimore, USA).

Management team: Sam Gledhill (Movember Foundation, Melbourne, Australia), Mark Buzza (Movember Foundation, Melbourne, Australia), Chris Bangma (Erasmus Medical Center, Rotterdam, The Netherlands), Monique Roobol (Erasmus Medical Center, Rotterdam, The Netherlands), Sophie Bruinsma / Jozien Helleman (Erasmus Medical Center, Rotterdam, The Netherlands).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.eururo.2019.07.041.

References

[1] Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. Eur Urol 2014;65:1046–55.

[2] Tosoian JJ, Carter HB, Lepor A, Loeb S. Active surveillance for prostate cancer: current evidence and contemporary state of practice. Nat Rev Urol 2016;13:205–15.

[3] Drost FH, Rannikko A, Valdagni R, et al. Can active surveillance really reduce the harms of overdiagnosing prostate cancer? A reflection of real life clinical practice in the PRIAS study. Transl Androl Urol 2018;7:98–105.

[4] Bruinsma SM, Bangma CH, Carroll PR, et al. Active surveillance for prostate cancer: a narrative review of clinical guidelines. Nat Rev Urol 2016;13:151–67.

[5] Bokhorst LP, Valdagni R, Rannikko A, et al. A decade of active surveillance in the PRIAS study: an update and evaluation of the criteria used to recommend a switch to active treatment. Eur Urol 2016;70:954–60.

[6] Borghesi M, Ahmed H, Nam R, et al. Complications after systematic, random, and image-guided prostate biopsy. Eur Urol 2017;71:353–65.

[7] Bokhorst LP, Alberts AR, Rannikko A, et al. Compliance rates with the Prostate Cancer research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. Eur Urol 2015;68:814–21.

[8] Luckenbaugh AN, Auffenberg GB, Hawken SR, et al. Variation in guideline concordant active surveillance followup in diverse urology practices. J Urol 2017;197:621–6.

[9] Bokhorst LP, Lepisto I, Kakehi Y, et al. Complications after prostate biopsies in men on active surveillance and its effect on receiving further biopsies in the Prostate cancer Research International: active Surveillance (PRIAS) study. BJU Int 2016;118:366–71.

[10] Ankerst DP, Xia J, Thompson Jr IM, et al. Precision medicine in active surveillance for prostate cancer: development of the Canary-early Detection Research Network active surveillance biopsy risk calculator. Eur Urol 2015;68:1083–8.

[11] Lin DW, Newcomb LF, Brown MD, et al. Evaluating the four kallikrein panel of the 4Kscore for prediction of high-grade prostate cancer in men in the Canary Prostate Active Surveillance Study. Eur Urol 2017;72:448–54.

[12] Bruinsma SM, Zhang L, Roobol MJ, et al. The Movember Foundation's GAP3 cohort: a profile of the largest global prostate cancer active surveillance database to date. BJU Int 2018;121:737–44.

[13] Newcomb LF, Brooks JD, Carroll PR, et al. Canary Prostate Active Surveillance Study: design of a multi-institutional active surveillance cohort and biorepository. Urology 2010;75:407–13.

[14] Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185–90.

[15] Ma TM, Tosoian JJ, Schaeffer EM, et al. The role of multiparametric magnetic resonance imaging/ultrasound fusion biopsy in active surveillance. Eur Urol 2017;71:174–80.

[16] Ehdaie B, Vertosick E, Spaliviero M, et al. The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance. J Urol 2014;191:660–4.

[17] Adamy A, Yee DS, Matsushita K, et al. Role of prostate specific antigen and immediate confirmatory biopsy in predicting progression during active surveillance for low risk prostate cancer. J Urol 2011;185:477–82.

[18] Satasivam P, Poon BY, Ehdaie B, Vickers AJ, Eastham JA. Can confirmatory biopsy be omitted in patients with prostate cancer favorable diagnostic features on active surveillance? J Urol 2016;195:74–9.

[19] Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol 2015;33:272–7.

[20] Cooperberg MR, Cowan JE, Hilton JF, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol 2011;29:228–34.

[21] Tran GN, Leapman MS, Nguyen HG, et al. Magnetic resonance imaging-ultrasound fusion biopsy during prostate cancer active surveillance. Eur Urol 2017;72:275–81.
[22] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38.
[23] Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015;68:279–89.
[24] Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 2016;352:i6.
[25] Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med 2004;23:2567–86.
[26] Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167–76.
[27] Loeb S, Tosoian JJ. Biomarkers in active surveillance. Transl Androl Urol 2018;7:155–9.
[28] Schoots IG, Petrides N, Giganti F, et al. Magnetic resonance imaging in active surveillance of prostate cancer: a systematic review. Eur Urol 2015;67:627–36.
[29] Borque-Fernando A, Rubio-Briones J, Esteban LM, et al. The management of active surveillance in prostate cancer: validation of the Canary Prostate Active Surveillance Study risk calculator with the Spanish Urological Association Registry. Oncotarget 2017;8:108451–6.
[30] Nieboer D, van der Ploeg T, Steyerberg EW. Assessing discriminative performance at external validation of clinical prediction models. PLoS One 2016;11:e0148820.
[31] Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol 2018;74:796–804.
[32] Roobol MJ, Verbeek J, Nieboer D, et al. Adherence to active surveillance protocols for low-risk prostate cancer: results of the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) initiative. Eur Urol 2017;16:e2632–5.