BJU International

Predicting biochemical recurrence and prostate cancer-specific mortality after radical prostatectomy: comparison of six prediction models in a cohort of patients with screening- and clinically detected prostate cancer

Sebastiaan Remmers*, Jan F. M. Verbeek*, Daan Nieboer*†, Theo van der Kwast‡§ and Monique J. Roobol*

*Department of Urology , †Department of Public Health , ‡Department of Pathology, Erasmus University Medical Centre, Rotterdam, The Netherlands , and §Department of Pathology, Toronto General Hospital, Toronto, ON, Canada

Objectives

To perform a comparison and external validation of three models predicting biochemical recurrence (BCR) and three models predicting prostate cancer (PCa)-specific mortality (PCSM) in a screening setting, i.e. patients with screening-detected PCa (S-PCa) and in those with clinically detected PCa (C-PCa).

Subjects and Methods

We retrospectively evaluated 795 men with S-PCa, from the European Randomized Study of Screening for Prostate Cancer, Rotterdam, and 1123 men with C-PCa initially treated with RP. The discriminative ability of the models was assessed according to the area under the curve (AUC) of the receiver-operating characteristic, and calibration was assessed graphically using calibration plots.

Results

The median (interquartile range [IQR]) follow-up for the S-PCa group was 10.4 (6.8–14.3) years and for the C-PCa group it was 8.8 (4.8–12.9) years. A total of 123 men with S-PCa (15%) and 389 men with C-PCa (35%) experienced BCR. Of the men with S-PCa and BCR, 24 (20%) died from PCa and 29 (23%) died from other causes. Of the men with C-PCa and BCR, 68 (17%) died from PCa and 105 (27%) died from other causes. The discrimination of the models predicting BCR or PCSM was higher for men with S-PCa (AUC: BCR 0.77–0.84, PCSM 0.60–0.77) than for the men with C-PCa (AUC: BCR 0.75–0.79, PCSM 0.51–0.68) as a result of the similar patient characteristics of the men with S-PCa in the present study and those of the cohorts used to develop these models. The risk of BCR was typically overestimated, while the risk of PCSM was typically underestimated.

Conclusion

Prediction models for BCR showed good discrimination and reasonable calibration for both men with S-PCa and men with C-PCa, and even better discrimination for men with S-PCa. For PCSM, the evaluated models are not applicable in both settings of this Dutch cohort as a result of substantial miscalibration. This warrants caution when using these models to communicate future risks in other clinical settings.

Keywords

clinical decision-making, nomograms, probability, prostatectomy, prostatic neoplasms, #ProstateCancer, #PCSM

Introduction

Radical prostatectomy (RP) is a surgical treatment used for localized prostate cancer (PCa), with overall high disease-specific survival rates. It has been demonstrated that the 8-year disease-specific survival for patients with pathological Gleason score of ≥4 + 4 is > 80% [1,2]. RP is considered successful if the PSA level after surgery drops to < 0.1 ng/mL. In 10–20% of patients, PSA level will increase again within a median period of 10 years after the RP procedure [3,4]. This biochemical recurrence (BCR) is defined as two successive PSA levels ≥0.2 ng/mL and can eventually lead to PCa-specific mortality (PCSM) in 2.5–17% of patients [3,5].

As such, early detection of BCR is important for management of the disease [6]. To aid in clinical decision-making,

BJU International | doi:10.1111/bju.14790
Published by John Wiley & Sons Ltd on behalf of BJU International. www.bjui.org

multivariable prediction models have been developed to provide the individual probability of BCR after RP, or PCSM following BCR after RP, estimated on the basis of patient and tumour characteristics. The estimates of these models are often included in decision-making without sufficient knowledge of external validity, which potentially could lead to erroneous individual risk assessments, thereby threatening adequate decision-making regarding, for example, initiation of salvage treatment.

We identified six prediction models, three of which were for predicting BCR after RP: the model by Walz et al. [7] (2-year BCR); the CAPRA nomogram by Cooperberg et al. [8] (5-year BCR); and the Memorial Sloan-Kettering Cancer Center (MSKCC) nomogram (5-year BCR; http://www.mskcc.org/cancer-care/adult/prostate/prediction-tools; updated on 4 October 2018). The remaining three models were for predicting PCSM following BCR after RP: models by Dell'Oglio et al. [9] (5-year PCSM); Brockman et al. [10] (15-year PCSM); and Eggener et al. [11] (15-year PCSM). All cohorts underlying these models consisted of referral-based patients. It was previously reported by Loeb et al. [3] that patients with screening-detected PCa (S-PCa) had better progression-free survival compared to men with clinically detected PCa (C-PCa) (i.e. PCa detected in clinical practice). Although less prominent, this was also true for PCSM.

The aims of the present study were to assess the predictive performance (discrimination and calibration) of the six prediction models in a true population-based PSA-driven screening setting, i.e. patients with S-PCa within the Dutch part of European Randomized Study of Screening for Prostate Cancer (ERSPC) [12], and to compare the performance of the model in a cohort of men with C-PCa from the same region.

Subjects and Methods

The ERSPC is a randomized controlled trial which studies the effect of PSA-based screening on PCa mortality. Randomization of men in the Dutch arm of the study was initiated in December 1993. The ERSPC study characteristics have been described previously [12]. In short, in the Dutch arm of ERSPC a total of 42 376 men were randomized to either a screening arm or a control arm. In the screening arm, men were re-screened every 4 years and received a PSA test and TRUS-guided prostate biopsy if PSA was $\geq$ 3.0 ng/mL. In the early years an abnormal DRE could trigger biopsy if PSA was < 4.0 ng/mL. Screening was discontinued if the patient was diagnosed with PCa or reached the age of 74 years. In the control arm, men received no active screening; treatment was offered if the patient was diagnosed with PCa after referral.

In addition to data on the men with C-PCa in the ERSPC Rotterdam study, data were available from men with C-PCa initially treated with RP in the Erasmus Medical Centre PCa study database. These men were not randomized as part of the ERSPC trial, but their PCa was clinically detected in the same region of Rotterdam. For both patients with S-PCa and C-PCa, treatment was carried out according to contemporary clinical practice.

After RP, pathological T-stage, nodal stage, Gleason grade, degree of extraprostatic extension, and surgical margins were assessed based on TNM 1992 classification. In addition, PSA was regularly measured after RP, and events of disease progression and death were recorded through semi-annual chart review. Cause of death was assessed by either a Cause of Death Committee according to a fixed algorithm if men were included in the ERSPC Rotterdam study [13] or was based on chart review by experienced urologists if men were not included in the ERPSC Rotterdam study. For the BCR prediction models, the aggregated patient criteria of the cohorts of the studies in which the models were developed (i.e. the development cohorts) comprised absence of neoadjuvant and adjuvant therapy. For the PCSM prediction models, the aggregated patient criteria of the original cohorts comprised presence of BCR, and men with or without adjuvant therapy were included. For all prediction models, men were excluded if they had pathological T0 stage, had unknown surgery data, had an unfinished surgery, had a diagnosis of PCa prior to randomization or had undergone cystectomy.

Statistical Analyses

The probability of experiencing BCR or PCSM was calculated using the required information for the prediction model under study (Table 1 and Table S1). Comparison was performed on the basis of discrimination and calibration. Discrimination refers to the ability of a prediction model to discriminate between patients with and without the event of interest and was quantified using the time-dependent area under the of the curve (AUC) of the receiver-operating characteristic, with nearest-neighbour estimation [14,15]. Calibration refers to the agreement between predicted probabilities and observed probabilities and was quantified using the calibration slope. Ideally, the calibration slope is equal to one. If the calibration slope is more than one, there will be underestimation of the predicted risk. If the calibration slope is less than one, there will be overestimation of the predicted risk. For the calibration plots, patients were divided into quintiles based on predicted event probability. The average predicted event probability of each subset of quintiles was plotted against the Kaplan–Meier event probability of each corresponding subset of quintiles [16].

For the models predicting BCR, patients without progression were censored at the last recorded date of PSA measurement [17]. All statistical analyses were performed using R version 3.5.1, survival analyses were conducted with R-package survival [18], missing values were imputed five times with R-

Table 1 Overview of the included predictors per study for progression or prostate cancer-specific mortality in the prediction model

	Age at diagnosis	Age at BCR	Time to BCR	PSA for RP	PSA after RP	pGS	pT	pN	EPE	SM	SVI	RT	HT
BCR													
Walz et al. [7]					+	+		+	+	+	+		
Cooperberg et al. [8]				+		+		+	+	+	+		
MSKCC updated on 4 October 2018					+	+			+	+	+		
PCSM													
Dell'Oglio et al. [9]		+	+			+	+	+		+		+	+
Brockman et al. [10]	+		+	+		+		+	+	+	+		
Eggener et al. [11]	+			+		+		+	+	+	+		

BCR, biochemical recurrence; ECE, extraprostatic extension; HT, hormone therapy; MSKCC, Memorial Sloan-Kettering Cancer Center; OC, organ-confined; PCSM, prostate cancer-specific mortality; pGS, pathological Gleason score; pM, pathological M-stage; pN, pathological N-stage; pT, pathological T-stage; RT, radiotherapy; SM, surgical margins; SVI, seminal vesicle invasion.

package mice [19], receiver-operating characteristic curves were computed by R-package survivalROC [20], and visualization of the plots was performed with R-package ggplot2 [21].

Results

Biochemical Recurrence after Radical Prostatectomy

For the models predicting BCR after RP, the patient cohorts consisted of 795 men with S-PCa who underwent surgery between January 1994 and June 2016 (median November 1999), and 1123 men with C-PCa who underwent surgery between February 1977 and August 2016 (median February 2001; Fig. 1). The overall median (interquartile range [IQR]) age at time of RP of the men with S-PCa was 66 (62–69) years and for men with C-PCa it was 64 (59–68) years (Table 2). The median (IQR) preoperative PSA level for men with S-PCa was 4.9 (3.6–7.7) ng/mL and for the men with C-PCa it was 8.9 (5.7–14.7) ng/mL. A total of 184 men (23%) in the S-PCa group and 535 men (47%) in the C-PCa group had a pathological T-stage of ≥T3. The overall median (IQR) follow-up time for men with S-PCa was 10.4 (6.8–14.3) years and for men with C-PCa it was 8.8 (4.8–12.9) years. A total of 123 men with S-PCa (15%) and 389 men with C-PCa (35%) experienced BCR. The overall median (IQR) age at the

Fig. 1 Flowchart of the patient inclusion process for the external validation of the models predicting biochemical recurrence (BCR) after radical prostatectomy. PCa, prostate cancer.

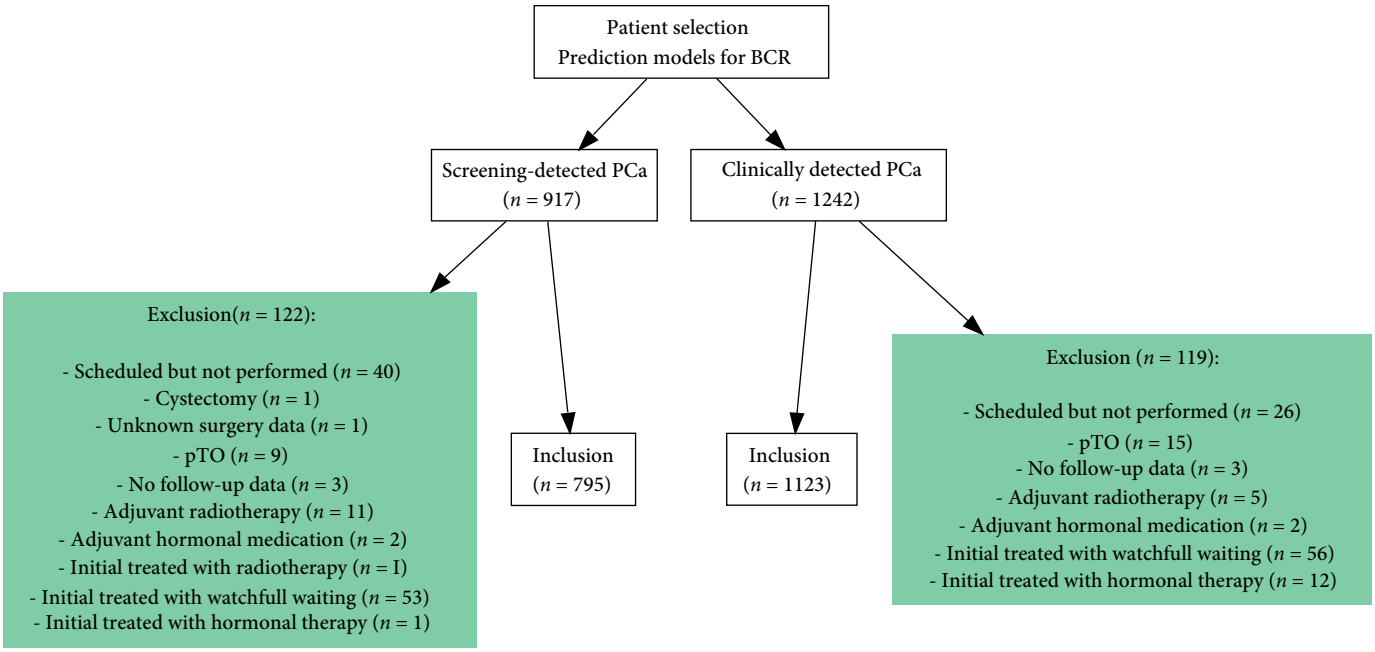
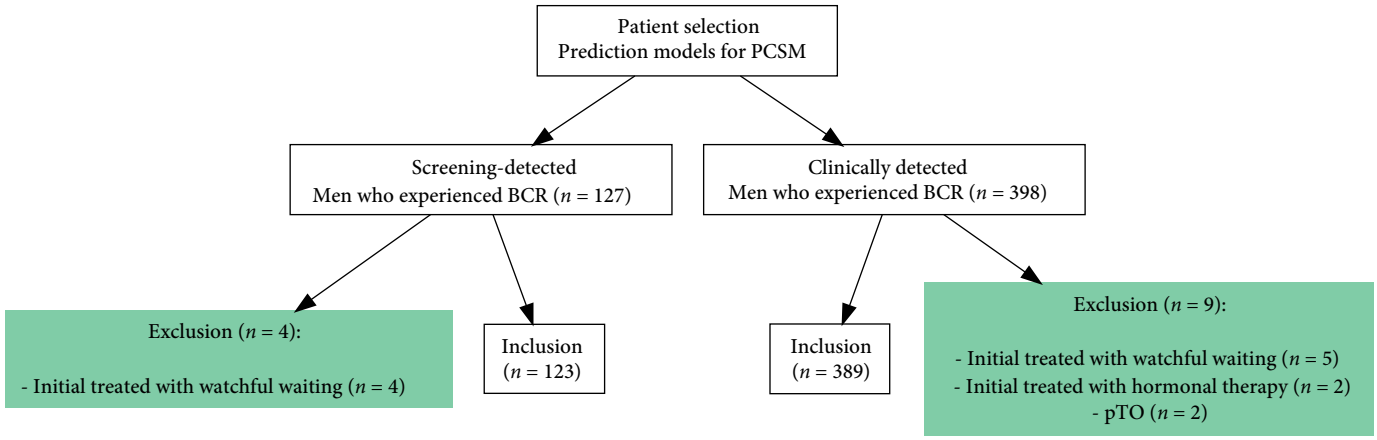


Table 2 Baseline characteristics of the patients who were treated by radical prostatectomy within the available follow-up time

	Overall all patients (*n* = 1918)			BCR (*n* = 512)			*P**
	Overall (*n* = 1918)	S-PCa (*n* = 795)	C-PCa (*n* = 1123)	Overall (*n* = 512)	S-PCa (*n* = 123)	C-PCa (*n* = 389)	
Age at surgery, years							
Median (IQR)	64.9 (60.6–68.5)	65.9 (62.1–69.0)	64.0 (59.3–68.1)	65.0 (60.3—68.4)	66.9 (63.2–68.9)	63.9 (59.4–68.2)	<0.001
Preoperative PSA, ng/mL							
Median (IQR)	6.8 (4.4–11.2)	4.9 (3.6–7.7)	8.9 (5.9–14.7)	10.2 (6.0–17.5)	6.5 (5.1–12.1)	11.9 (6.9–19.0)	<0.001
Pathological stage, *n* (%)							
T2A–T2B	395 (21)	190 (24)	205 (18)	58 (11)	14 (11)	44 (11)	<0.001
T2C	804 (42)	421 (53)	383 (34)	103 (20)	42 (34)	61 (16)	
T3A–T3C	625 (33)	160 (20)	465 (41)	290 (57)	56 (46)	234 (60)	
T4A–T4B	94 (5)	24 (3)	70 (6)	61 (12)	11 (9)	50 (13)	
Pathological Gleason score, *n* (%)							
≤3 + 3	676 (35)	422 (53)	254 (23)	62 (12)	25 (20)	37 (10)	<0.001
3 + 4	543 (28)	228 (29)	315 (28)	137 (27)	38 (31)	99 (25)	
4 + 3	125 (7)	49 (6)	76 (7)	58 (11)	22 (18)	36 (9)	
≥4 + 4	160 (8)	52 (7)	108 (10)	75 (15)	28 (23)	47 (12)	
Unknown	414 (22)	44 (6)	370 (33)	180 (35)	10 (8)	170 (44)	
Extraprostatic extension, *n* (%)							
None	1023 (53)	315 (40)	708 (63)	236 (46)	24 (20)	212 (54)	<0.001
Internal	309 (16)	265 (33)	44 (4)	44 (9)	35 (28)	9 (2)	
External	477 (25)	142 (18)	335 (30)	200 (39)	49 (40)	151 (39)	
Unknown	109 (6)	73 (9)	36 (3)	32 (6)	15 (12)	17 (4)	
Surgical margins, *n* (%)							
Negative	956 (50)	590 (74)	366 (33)	115 (22)	55 (45)	60 (15)	<0.001
Positive	631 (33)	194 (24)	437 (39)	290 (57)	67 (54)	223 (57)	
Unknown	331 (17)	11 (1)	320 (28)	107 (21)	1 (1)	106 (27)	
Seminal vesicle invasion, *n* (%)							
Negative	1664 (87)	741 (93)	923 (82)	349 (68)	95 (77)	254 (65)	<0.001
Positive	219 (11)	39 (5)	180 (16)	149 (29)	26 (21)	123 (32)	
Unknown	35 (2)	15 (2)	20 (2)	14 (3)	2 (2)	12 (3)	
Nodal stage, *n* (%)							
NX	240 (13)	163 (21)	77 (7)	24 (5)	11 (9)	13 (3)	<0.001
N0	1474 (77)	626 (79)	848 (76)	420 (82)	109 (89)	311 (80)	
N1	63 (3)	6 (1)	57 (5)	48 (9)	3 (2)	45 (12)	
Unknown	141 (7)	0 (0)	141 (13)	20 (4)	0 (0)	20 (5)	
Time till BCR, years							
Median (IQR)	–	–	–	2.5 (1.1–5.5)	3.3 (1.7–7.3)	2.4 (1.0–5.1)	<0.001
Follow-up time after RP, years							
Median (IQR)	9.7 (5.4 –13.6)	10.4 (6.8–14.3)	8.8 (4.8–12.9)	11.0 (6.8–15.2)	12.3 (8.0–15.9)	10.4 (6.7–14.5)	0

*BCR, biochemical recurrence; C-PCa, clinically detected prostate cancer; IQR, interquartile range; RP, radical prostatectomy; S-PCa, screening-detected prostate cancer. *Comparing characteristics of men experiencing BCR for men with S-PCa and C-PCa. Continuous variables were analysed using the Mann–Whitney U-test; categorical variables were analysed using the chi-squared test.*

Fig. 2 Flowchart of the patient inclusion for the external validation of the models predicting prostate cancer-specific mortality (PCSM) following biochemical recurrence (BCR) after radical prostatectomy.



time of RP of the men who experienced BCR was 67 (63–69) years for S-PCa group and 64 (59–68) for the C-PCa group.

There was no substantial difference in discrimination between the models, but a higher AUC was found for men with S-PCa (AUC range 0.77–0.84) vs men with C-PCa (AUC range 0.75–0.79; Fig. S1). For the model by Cooperberg et al. [8], the calibration slope was 0.79 for C-PCa and 1.003 for S-PCa, for the model by Walz et al. [7] it was 0.70 for C-PCa and 0.92 for S-PCa, and for the MSKCC model it was 0.25 for C-PCa and 0.66 for S-PCa. The calibration of the models was in general more accurate for the men with C-PCa: the degree of overestimation of the predicted probabilities of experiencing BCR was higher for the men with S-PCa (Fig. S2) as the calibration slope was < 45 degrees. The calibration of the MKSCC model for both S-PCa and C-PCa is shown in detail in Fig. S3.

Prostate Cancer-Specific Mortality Following Biochemical Recurrence after Radical Prostatectomy

A total of 123 men with S-PCa and 389 men with C-PCa experienced BCR (Fig. 2). The median (IQR) time to BCR after diagnosis was 3.3 (1.7–7.3) years for men with S-PCa and 2.4 (1.0–5.1) years for men with C-PCa (Table 2). For the men in both the S-PCa and C-PCa groups who experienced BCR, about the same proportion of men died from either PCa or other causes. To elaborate, of the men with S-PCa who experienced BCR, 24 (20%) died from PCa and 29 (24%) died from other causes. Of the men with C-PCa who experienced BCR, 68 (17%) died from PCa and 105 (27%) died from other causes (Table 3). Among the men who died from PCa, the median (IQR) follow-up in the S-PCa group was 7.2 (4.9–11.8) years, and in the C-PCa group it was 8.9 (6.3–11.2) years.

The discrimination of the models predicting PCSM for the men with S-PCa was between 0.60 and 0.77, and for the men with C-PCa it was between 0.51 and 0.68 (Fig. S4). The calibration of the models in men with S-PCa was more accurate than in men with C-PCa (Fig. S5). For the model by Brockman et al. [10], the calibration slope was 0.08 for the C-PCa and 0.84 for S-PCa group, for the model by Eggener et al. [11] it was 0.29 for the C-PCa and 0.43 for the S-PCa group, and for the model by Dell'Oglio et al. [9] it was 0.43 for the C-PCa and 0.92 for the S-PCa group. In addition, in all models there was substantial miscalibration regarding the probability of dying from PCa for both men with S-PCa and those with C-PCa.

Discussion

In the present study, we performed an external validation of models predicting BCR or PCSM. For the models predicting BCR, we found that the difference in their discriminative ability for the men with S-PCa and for those with C-PCa was minimal, but was higher for men with S-PCa. The calibration plots, however, showed a higher degree of overestimation of the probability of experiencing BCR for the men with S-PCa. This overestimation of experiencing BCR is probably attributable to differences between a population-based screening setting as compared to a referral population, where the first leads to more low-risk patients [22,23]. In addition, independent of setting, early diagnosis coincides with better prognosis [23].

The models with the best calibration slope were those by Cooperberg et al. [8] and Walz et al. [7] for predicting BCR. These models are most suitable for clinical utility, taking into account that the baseline hazard needs to be modified. This is probably the result of regional differences. Difference in discrimination between the development cohort and the validation cohort are likely to reflect differences in case mix if the parameter estimates are correct [24].

Table 3 Baseline characteristics of the patients who experienced biochemical recurrence and died within the available follow-up time

	Overall mortality (*n* = 226)			PCSM (*n* = 92)			Other cause of death (*n* = 134)		
	Overall (*n* = 226)	S-PCa (*n* = 53	C-PCa (*n* = 173)	Overall (*n* = 92)	S-PCa (*n* = 24)	C-PCa (*n* = 68)	Overall (*n* = 134)	S-PCa (*n* = 29)	C-PCa (*n* = 105)
Age at surgery, years									
Median (IQR)	65.2 (61.3–68.4)	67.0 (63.0–69.0)	64.3 (60.2–68.1)	63.0 (60.4–67.6)	66.0 (62.3–68.7)	62.8 (59.2–67.2)	65.6 (62.4–68.9)	67.7 (65.3–69.2)	65.2 (61.3–68.8)
Preoperative PSA, ng/mL									
Median (IQR)	10.5 (5.8–18.7)	9.6 (5.0–10.9)	12.4 (7.4–21.4)	10.4 (6.0–21.6)	7.4 (5.8–12.3)	15.3 (6.2–24.5)	10.5 (5.7–16.5)	6.3 (5.0–9.2)	12.3 (7.6–18.8)
Pathological stage, *n* (%)									
T2A–T2B	29 (13)	6 (11)	23 (13)	6 (7)	2 (8)	4 (6)	23 (17)	4 (14)	19 (18)
T2C	31 (14)	20 (38)	11 (6)	8 (9)	6 (25)	2 (3)	23 (17)	14 (48)	9 (9)
T3A–T3C	133 (59)	22 (42)	111 (64)	49 (53)	11 (46)	38 (56)	73 (54)	11 (38)	62 (59)
T4A–T4B	33 (15)	5 (9)	28 (16)	18 (20)	5 (21)	13 (19)	15 (11)	0 (0)	15 (14)
Pathological Gleason score, *n* (%)									
≤3 + 3	23 (10)	10 (19)	13 (8)	4 (4)	2 (8)	2 (3)	19 (14)	8 (28)	11 (10)
3 + 4	45 (20)	16 (30)	29 (17)	19 (21)	5 (21)	14 (21)	26 (19)	11 (38)	15 (14)
4 + 3	26 (12)	10 (19)	16 (9)	15 (16)	6 (25)	9 (13)	11 (8)	4 (14)	7 (7)
≥4 + 4	28 (12)	13 (25)	15 (9)	17 (18)	12 (50)	5 (7)	12 (9)	2 (7)	10 (10)
Unknown	104 (46)	4 (8)	100 (58)	38 (41)	0 (0)	38 (56)	66 (49)	4 (14)	62 (59)
Extraprostatic extension, *n* (%)									
None	122 (54)	9 (17)	113 (65)	55 (60)	6 (25)	49 (72)	67 (50)	3 (10)	64 (61)
Internal	16 (7)	14 (26)	2 (1)	3 (3)	3 (13)	0 (0)	13 (10)	11 (38)	2 (2)
External	76 (34)	24 (45)	52 (30)	31 (34)	13 (54)	18 (26)	45 (34)	11 (38)	34 (32)
Unknown	12 (5)	6 (11)	6 (3)	3 (3)	2 (8)	1 (1)	9 (7)	4 (14)	5 (5)
Surgical margins, *n* (%)									
Positive	125 (55)	29 (55)	96 (55)	59 (64)	13 (54)	46 (68)	66 (49)	16 (55)	50 (48)
Negative	37 (16)	24 (45)	13 (8)	15 (16)	11 (46)	4 (6)	22 (16)	13 (45)	9 (9)
Unknown	64 (28)	0 (0)	64 (37)	18 (20)	0 (0)	18 (26)	0 (0)	0 (0)	0 (0)
Seminal vesicle invasion, *n* (%)									
Positive	81 (36)	12 (23)	69 (40)	50 (54)	9 (38)	41 (60)	31 (23)	3 (10)	28 (27)
Negative	141 (62)	41 (77)	100 (58)	42 (46)	15 (63)	27 (40)	99 (74)	26 (90)	73 (70)
Unknown	4 (2)	0 (0)	4 (2)	0 (0)	0 (0)	0 (0)	4 (3)	0 (0)	4 (4)
Nodal stage, *n* (%)									
NX	4 (2)	4 (8)	0 (0)	3 (3)	3 (13)	0 (0)	1 (1)	1 (3)	0 (0)
N0	189 (84)	47 (89)	142 (82)	73 (79)	19 (79)	54 (79)	116 (87)	28 (97)	88 (84)
N1	31 (14)	2 (4)	29 (17)	16 (17)	2 (8)	14 (21)	15 (11)	0 (0)	15 (14)
Unknown	2 (1)	0 (0)	2 (1)	0 (0)	0 (0)	0 (0)	2 (1)	0 (0)	2 (2)
Follow-up time after RP, years									
Median (IQR)	9.9 (6.4–13.6)	10.5 (6.0–14.1)	9.8 (6.5–13.3)	8.7 (6.1–11.4)	7.2 (4.9–11.8)	8.9 (6.3–11.2)	11.0 (6.5–14.3)	13.0 (9.7–14.6)	9.8 (6.4–14.0)

BCR, biochemical recurrence; C-PCa, clinically detected prostate cancer; IQR, interquartile range; PCSM, prostate cancer-specific mortality; RP, radical prostatectomy; S-PCa, screening-detected prostate cancer.

Discrimination will decrease if the case mix of the validation cohort is homogeneous with the development cohort (e.g. patient inclusion is based on stringent criteria). To elaborate, the IQR of the preoperative PSA level in the model of Brockman et al. [10] was 5.6–13.4 ng/mL, which is closer to the PSA level of men with S-PCa (i.e. 5.1–17.5 ng/mL) as compared to those with C-PCa (i.e. 6.9–19.9 ng/mL). In addition, the calibration of the models predicting PCSM for C-PCa and S-PCa was far from optimal, but slightly better for the latter. It must be noted that the available follow-up data for the development of the models of Brockman et al. [10] and Eggener et al. [11] might not have been sufficient to provide reliable predictions on 15-year PCSM: the median (IQR) follow-up of patients in the cohort described by Brockman et al. [10] was 7 (4.0–10.8) years and in the cohort described by Eggener et al. [11] it was only 4.7 (2.0–7.8) years.

Some of the prediction models used in the present study have been externally validated previously. For example, Lughezzani et al. [25] validated the CAPRA prediction model of Cooperberg et al. [8], and Tanaka et al. [26] validated the prediction model of the September 2013 MSKCC nomogram. Their results are in line with our study; however, the difference between these external validations and the present one is that we also included men with PCa detected purely on the basis of an elevated PSA level and this way we clearly show that differences in setting can affect the performance of a prediction model.

The present external validation study is not without potential limitations. The event rate for PCSM was limited as, within the available follow-up, only 92 men with BCR (18%) died from PCa after RP. In addition, there was no central review of the RP specimen. Nevertheless, this reflects clinical practice and it is unlikely that this would have affected the results. In addition, in the present study we could not identify the best performing model as the models differed in endpoints with regard to timeline of predictions for both BCR and PCSM; for BCR predictions were made for 2 and 5 years, and for PCSM predictions were made for 5 and 15 years.

The strength of the present study is that the cause of death of the men who participated in the ERSPC Rotterdam was evaluated by a dedicated Cause of Death Committee. This committee collected the complete medical record to obtain information about metastases and progression, and whether progressive metastatic disease caused death was decided by means of a flow chart. Without sufficient information to establish a cause of death, information was linked to Statistics Netherlands to analyse death certificates. By using a Cause of Death Committee, we could standardize the cause of death of the men who participated in the ERSPC Rotterdam study. The cause of death of the men who participated in the Erasmus Medical Centre prostate cancer study database were evaluated based on chart review by experienced urologists.

Van den Broeck et al. [27] performed a recent meta-analysis of men with curative treatment for PCa and who experienced BCR in order to determine whether BCR was associated with oncological outcomes and to identify prognostic factors for oncological outcomes after BCR. They identified that, for men primarily treated with RP, a PSA doubling time of > 1 year and a pathological Gleason score < 8 indicated low risk for distant metastasis, PCSM and overall mortality. A PSA doubling time of ≤ 1 year or a pathological Gleason score ≥ 8 indicated high risk.

From a clinical perspective, our results have important implications. While physicians and also patients are encouraged to use the information from the prediction model in their shared decision-making process, they should also acknowledge the potential limitations of a prediction model and when choosing a certain prediction model they should consider discrepancies between their clinical setting and the setting in which the model was developed.

In conclusion, we performed an external validation of models predicting either BCR or PCSM after RP in a PSA-based screening setting and a more clinical setting, including patients after referral. Prediction models for BCR after RP showed good discrimination and reasonable calibration in both settings; however, when using predictions related to PCSM after RP, prudence in their use and recalibration is advised given their very limited performance in this external validation.

Acknowledgements

The Dutch part of ERSPC was funded by the Dutch Cancer Society (KWF 94-869, 98-1657, 2002-277, 2006-3518, 2010-4800); The Netherlands Organization for Health Research and Development (ZonMW-002822820, 22000106, 50-50110-98-311, 62300035), The Dutch Cancer Research Foundation (SWOP), and an unconditional grant from Beckman-Coulter-Hybritech Inc. Special thanks are extended to Dr Mark F. Wildhagen for providing additional RP data from the Erasmus Medical Centre prostate cancer study database.

Conflicts of Interest

None declared.

References

1 **Kweldam CF**, **Wildhagen MF**, **Bangma CH**, **Leenders GJLH**. Disease-specific death and metastasis do not occur in patients with Gleason score ≤ 6 at radical prostatectomy. *BJU Int* 2015; 116: 230–5

2 **Pokala N**, **Trulson JJ**, **Islam M**. Long-term outcome following radical prostatectomy for Gleason 8–10 prostatic adenocarcinoma. *World J Urol* 2014; 32: 1385–92

3 **Loeb S**, **Zhu X**, **Schroder FH**, **Roobol MJ**. Long-term radical prostatectomy outcomes among participants from the European

Randomized Study of Screening for Prostate Cancer (ERSPC) Rotterdam. *BJU Int* 2012; 110: 1678–83
4 **Han M, Partin AW, Zahurak M, Piantadosi S, Epstein JI, Walsh PC.** Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J Urol* 2003; 169: 517–23
5 **Freedland SJ, Humphreys EB, Mangold LA et al.** Risk of prostate cancer–specific mortality following biochemical recurrence after radical prostatectomy. *JAMA* 2005; 294: 433–9
6 **Roberts WB, Han M.** Clinical significance and treatment of biochemical recurrence after definitive therapy for localized prostate cancer. *Surg Oncol* 2009; 18: 268–74
7 **Walz J, Chun FKH, Klein EA et al.** Nomogram predicting the probability of early recurrence after radical prostatectomy for prostate cancer. *J Urol* 2009; 181: 601–8
8 **Cooperberg MR, Hilton JF, Carroll PR.** The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer* 2011; 117: 5039–46
9 **Dell'Oglio P, Suardi N, Boorjian SA et al.** Predicting survival of men with recurrent prostate cancer after radical prostatectomy. *Eur J Cancer* 2016; 54: 27–34
10 **Brockman JA, Alanee S, Vickers AJ et al.** Nomogram predicting prostate cancer–specific mortality for men with biochemical recurrence after radical prostatectomy. *Eur Urol* 2015; 67: 1160–7
11 **Eggener SE, Scardino PT, Walsh PC et al.** Predicting 15-year prostate cancer specific mortality after radical prostatectomy. *J Urol* 2011; 185: 869–75
12 **Roobol MJ, Kirkels WJ, Schroder FH.** Features and preliminary results of the Dutch centre of the ERSPC (Rotterdam, the Netherlands). *BJU Int* 2003; 92(Suppl 2): 48–54
13 **Koning HJD, Blom J, Merkelbach JW et al.** Determining the cause of death in randomized screening trial(s) for prostate cancer. *BJU Int* 2003; 92: 71–8
14 **Blanche P, Kattan MW, Gerds TA.** The c-index is not proper for the evaluation of t-year predicted risks. *Biostatistics* 2019; 20; 347–57
15 **Heagerty PJ, Lumley T, Pepe MS.** Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* 2000; 56: 337–44
16 **Royston P.** Tools for checking calibration of a Cox model in external validation: approach based on individual event probabilities. *Stata Journal.* 2014; 14: 738–55
17 **Stish BJ, Pisansky TM, Harmsen WS et al.** Improved metastasis-free and survival outcomes with early salvage radiotherapy in men with detectable prostate-specific antigen after prostatectomy for prostate cancer. *J Clin Oncol* 2016; 34: 3864–71
18 **Therneau TM.** A package for survival analysis in S, 2015.
19 **Van Buuren S, Groothuis-Oudshoorn K.** mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011; 43: 1–67
20 **Heagerty PJ, Saha-Chaudhuri P.** survivalROC: Time-dependent ROC curve estimation from censored survival data, 2013.
21 **Wickham H.** ggplot2: Elegant Graphics for Data Analysis, 2009.
22 **Bokhorst LP, Venderbos LDF, Schröder FH, Bangma CH, Steyerberg EW, Roobol MJ.** Do treatment differences between arms affect the main outcome of ERSPC Rotterdam? *J Urol* 2015; 194: 336–42
23 **Bokhorst LP, Kranse R, Venderbos LDF et al.** Differences in treatment and outcome after treatment with curative intent in the screening and control arms of the ERSPC Rotterdam. *Eur Urol* 2015; 68: 179–82
24 **Nieboer D, van der Ploeg T, Steyerberg EW.** Assessing discriminative performance at external validation of clinical prediction models. *PLoS ONE* 2016; 11: e0148820
25 **Lughezzani G, Budäus 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
26 **Tanaka A, Ohori M, Paul L et al.** External validation of preoperative nomograms predicting biochemical recurrence after radical prostatectomy. *Jpn J Clin Oncol* 2013; 43: 1255–60
27 **Van den Broeck T, van den Bergh RCN, Arfi N et al.** Prognostic value of biochemical recurrence following treatment with curative intent for prostate cancer: a systematic review. *Eur Urol* 2018

Correspondence: Sebastiaan Remmers, Department of Urology, Erasmus University Medical Centre, P.O. Box 2040, Dr. Molewaterplein 40, 3000 CA, Rotterdam, The Netherlands.

e-mail: s.remmers@erasmusmc.nl

Abbreviations: BCR, biochemical recurrence; PCSM, prostate cancer-specific mortality; PCa, prostate cancer; S-PCa, screening-detected prostate cancer; C-PCa, clinically detected prostate cancer; IQR, interquartile range; MSKCC, Memorial Sloan-Kettering Cancer Center; ERSPC, Randomized Study of Screening for Prostate Cancer.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Discrimination of the models predicting BCR after RP.
Fig. S2. Calibration of models predicting BCR after RP.
Fig. S3. Kaplan-Meier with risk group as strata.
Fig. S4. Discrimination of the models predicting PCSM following BCR after RP.
Fig. S5. Calibration of models predicting PCSM following BCR after RP.
Table S1. Characteristics of the prediction models.