Predicting Prostate Cancer Death with Different Pretreatment Risk Stratification Tools: A Head-to-head Comparison in a Nationwide Cohort Study

(Article begins on next page)

20 February 2025

Predicting Prostate Cancer Death with Different Pretreatment Risk Stratification Tools: A Head-to-head Comparison in a Nationwide Cohort Study

Renata Zelic, Hans Garmo, Daniela Zugna, Pär Stattin, Lorenzo Richiardi Olof Akre, Andreas Pettersson

Abstract

Background

Numerous pretreatment risk classification tools are available for prostate cancer. Which tool is best in predicting prostate cancer death is unclear.

Objective

To systematically compare the prognostic performance of the most commonly used pretreatment risk stratification tools for prostate cancer.

Design, setting, and participants

A nationwide cohort study was conducted, including 154 811 men in Prostate Cancer data Base Sweden (PCBaSe) 4.0 diagnosed with nonmetastatic prostate cancer during 1998–2016 and followed through 2016.

Outcome measurements and statistical analysis

We compared the D'Amico, National Institute for Health and Care Excellence (NICE), European Association of Urology (EAU), Genito-Urinary Radiation Oncologists of Canada (GUROC), American Urological Association (AUA), National Comprehensive Cancer Network (NCCN), and Cambridge Prognostic Groups (CPG) risk group systems; the Cancer of the Prostate Risk Assessment (CAPRA) score; and the Memorial Sloan Kettering Cancer Center (MSKCC) nomogram in predicting prostate cancer death by estimating the concordance index (C-index) and the observed versus predicted cumulative incidences at different follow-up times.

Results and limitations

A total of 139 515 men were included in the main analysis, of whom 15 961 died from prostate cancer during follow-up. The C-index at 10 yr of follow-up ranged from 0.73 (95% confidence interval [CI]: 0.72–0.73) to 0.81 (95% CI: 0.80–0.81) across the compared tools. The MSKCC nomogram (C-index: 0.81, 95% CI: 0.80–0.81), CAPRA score (C-index: 0.80, 95% CI: 0.79–0.81), and CPG system (C-index: 0.78, 95% CI: 0.78–0.79) performed the best. The order of performance between the tools remained in analyses stratified by primary treatment and year of diagnosis. The predicted cumulative incidences were close to the observed ones, with some underestimation at 5 yr. It is a limitation that the study was conducted solely in a Swedish setting (ie, case mix).

Conclusions

The MSKCC nomogram, CAPRA score, and CPG risk grouping system performed better in discriminating prostate cancer death than the D'Amico and D'Amico-derived systems (NICE, GUROC, EAU, AUA, and NCCN). Use of these tools may improve clinical decision making.

Patient summary

There are numerous pretreatment risk classification tools that can aid treatment decision for prostate cancer. We systematically compared the prognostic performance of the most commonly used tools in a large cohort of Swedish men with prostate cancer. The Memorial Sloan Kettering Cancer Center nomogram, Cancer of the Prostate Risk Assessment score, and Cambridge Prognostic Groups performed best in predicting prostate cancer death. The use of these tools may improve treatment decisions.

1. Introduction

Given the wide variation of outcomes in men with prostate cancer [1], risk stratification is crucial for informed clinical decision making. The D'Amico risk stratification system, proposed in 1998, classifies patients into low-, intermediate-, and high-risk groups based on prostate-specific antigen (PSA) level, clinical tumor stage, and Gleason score at diagnosis [2], and it has become the main standard in clinical practice. However, several other risk stratification tools have been proposed, including risk grouping systems incorporating more granular clinicopathological information (eg, separating Gleason 3 + 4 from 4 + 3) or additional clinicopathological parameters (eg, extent of cancer in biopsy cores) [3], [4], [5], [6], risk scores [7], and nomograms [8].

Although the main purpose of the pretreatment risk stratification tools is to predict prostate cancer death in untreated men, most tools have used biochemical recurrence (BCR) rather than prostate cancer death as the endpoint and have been developed in radically treated rather than untreated men. Moreover, most tools have been developed in selected rather than population-based cohorts. To the best of our knowledge, no previous study has compared the most commonly used risk stratification tools head to head with respect to their ability to predict prostate cancer death.

Therefore, we used the Prostate Cancer data Base Sweden (PCBaSe), a population-based research database including both untreated and treated patients followed for prostate cancer death for up to 19 yr, to compare the prognostic performance of the following pretreatment risk stratification tools: the D'Amico [2], the National Institute for Health and Care Excellence (NICE) [9], the Genito-Urinary Radiation Oncologists of Canada (GUROC) [10], the American Urological Association (AUA) [11], the European Association of Urology (EAU) [12], the National Comprehensive Cancer Network (NCCN) [4], the Cambridge Prognostic Groups (CPG) [6], the Cancer of the Prostate Risk Assessment (CAPRA) score [7], and the Memorial Sloan Kettering Cancer Center (MSKCC) nomogram [13].

2. Patients and methods

2.1. Data sources

We used data from the PCBaSe version 4.0, a research database constructed by linkage between the National Prostate Cancer Register (NPCR) of Sweden and other population-based registers,

including the Total Population Register, National Patient Register, and Cause of Death Register [14], [15].

The NPCR is a clinical cancer register containing detailed data for >95% of all men diagnosed with prostate cancer in Sweden since 1998 [14], [16], including information on the date and hospital of diagnosis, mode of detection (PSA screening, lower urinary tract symptoms, and other symptoms), age, diagnostic PSA level, clinical tumor-node-metastasis stage, biopsy tumor differentiation, and planned primary treatment within 6 mo of diagnosis (deferred treatment [ie, active surveillance or watchful waiting], curative treatment [ie, radical prostatectomy or radiation therapy], or primary androgen deprivation therapy [ADT]). Since 2007, the NPCR contains information on prostate volume at diagnosis, total number of diagnostic biopsy cores, number of cores with cancer, total length of all biopsy cores, and combined length of cancer in all cores.

The Total Population Register contains date of death for virtually 100% and emigration for 91% of all Swedish citizens [17]. The National Patient Register includes in-patient medical and procedural discharge diagnoses according to International Classification of Disease (ICD) codes since 1987; ICD discharge codes up to 10 yr prior to the prostate cancer diagnosis were used to calculate the Charlson Comorbidity Index. The Cause of Death Register contains date, and underlying and contributory causes of death according to ICD-10 codes, with 86% agreement with the cause of death determined by medical record review for prostate cancer [18].

2.2. Study population

We included all men in PCBaSe 4.0 diagnosed with nonmetastatic (ie, not M1 or N1) prostate cancer between January 1, 1998 and December 31, 2016 ($n = 154\ 811$). The outcome was prostate cancer death, defined as prostate cancer listed as the underlying cause of death (ICD-10 code: C61). Date of emigration, and date and cause of death were available until December 31, 2016.

2.3. Statistical analyses

2.3.1. Missing data

The variables used in the different risk stratification tools are shown in Table 1, Table 2. Missing values for these variables were imputed using multiple imputation with chained equation [19], [20], with 50 imputations and 20 iterations per imputation. Information on cT2-cT3 substage (ie, cT2a, cT2b, cT2c, cT3a, and cT3b) is used in most risk stratification tools but is not recorded in the PCBaSe. We used a previously published cohort of men diagnosed with prostate cancer during 1995–2015, treated with proton-boost radiotherapy at the Uppsala University Hospital, Uppsala, Sweden [21], to predict cT2-cT3 substage for all patients in the PCBaSe. Details on the proportions of missing data, multiple imputation, and assignment of the cT2-cT3 substage are available in the Supplementary Methods and Supplementary Tables 1–7.

Table 1. Prostate cancer risk stratification criteria for different risk grouping systems.[a]

System	Very low risk	Low risk	Intermediate risk		High risk	Very high risk
			Favorable	Unfavorable		
D'Amico		PSA ≤ 10 and	PSA >10–20 or GS 7 or cT2b		PSA > 20 or GS 8–10 or cT2c	

System	Very low risk	Low risk	Intermediate risk		High risk	Very high risk
			Favorable	Unfavorable		
EAU		GS ≤6 and cT1c-2a PSA <10 and GS ≤6 (ISUP 1) and cT1c-2a	PSA 10–20 or GS 7 (ISUP 2–3) or cT2b		PSA > 20 or GS >7 (ISUP 4–5) or cT2c	
NICE		PSA <10 and GS ≤6 and cT1-2a	PSA 10–20 or GS 7 or cT2b		PSA > 20 or GS 8–10 or ≥cT2c	
GUROC		PSA ≤ 10 and GS ≤6 and cT1-2a	PSA ≤20 and GS ≤7 and cT1-2 not otherwise low-risk		PSA > 20 or GS 8–10 or ≥cT3a	
AUA	PSA <10 and ISUP 1 and cT1-2a and <34% positive cores and and PSAD <0.15	PSA <10 and ISUP 1 and cT1-2a	PSA 10–<20 or ISUP 2–3 or cT2b-2c		PSA ≥20 or ISUP 4–5 or ≥cT3	
AUA_i	PSA <10 and ISUP 1 and cT1-2a and <34% positive cores and and PSAD <0.15	PSA <10 and ISUP 1 and cT1-2a	ISUP 1 and PSA 10–<20 or ISUP 2 and PSA <10	ISUP 2 and (PSA 10–<20 or cT2b-2c) or ISUP 3 and PSA <20	PSA ≥20 or ISUP 4–5 or ≥cT3	
NCCN	PSA <10 and GS ≤6 (ISUP 1) and cT1c and	PSA <10 and GS ≤6 (ISUP 1) and cT1-2a	PSA 10–20 or GS 3 + 4 (ISUP 2) or cT2b-2c and <50%	PSA 10–20 or GS 3 + 4/4 + 3 (ISUP 2–3) or cT2b-2c	PSA > 20 or GS 4 + 4/4 + 5 (ISUP 4–5) or cT3a	GG1 5 or or cT3b-4

System	Very low risk	Low risk	Intermediate risk		High risk	Very high risk
			Favorable	Unfavorable		
	<3 positive cores and and PSAD <0.15		positive cores			
CPG		PSA <10 and GS 6 (ISUP 1) and cT1-T2	PSA 10–20 or GS 3 + 4 (ISUP 2) and cT1-T2	PSA 10–20 and GS 3 + 4 (ISUP 2) and cT1-T2 or GS 4 + 3 (ISUP 3) and cT1-T2	PSA > 20 or GS 8 (ISUP 4) or cT3	More than one of PSA > 20, GS 8 (ISUP 4), cT3 or GS 9–10 (ISUP 5) or cT4

AUA = American Urological Association; CPG = Cambridge Prognostic Groups; cT = clinical tumor stage; EAU = European Association of Urology; GS = Gleason score; GUROC = Genito-Urinary Radiation Oncologists of Canada; ISUP = International Society of Urological Pathology; NCCN = National Comprehensive Cancer Network; NICE = The National Institute for Health and Care Excellence; PCBaSe = Prostate Cancer data Base Sweden; PSA = prostate-specific antigen; PSAD = prostate-specific antigen density.

a

Information on the individual biopsy cores was not available in PCBaSe 4. Core level information (marked in red) could thus not be used in the construction of the risk groups.

Table 2. Variables included in the MSKCC nomograms and the CAPRA score.

	Nomograms		CAPRA score
	MSKCC	MSKCC_cores	
PSA			
Transformed (RCS)	×	×	
2.1–6.0			×
6.1–10.0			×
10.1–20.0			×
20.1–30.0			×
>30.0			×
Primary Gleason grade			
≤3 (3)	×	×	×
≥4 (4)	×	×	×
Secondary Gleason grade			
≤3 (3)	×	×	×
≥4 (4)	×	×	×

	Nomograms		CAPRA score
	MSKCC	MSKCC_cores	
Clinical tumor stage			
cT1	×	×	×
cT2			×
cT2a	×	×	
cT2b	×	×	
cT2c	×	×	
cT3			
cT3a			×
cT3b			
cT3+	×	×	
Age (yr)			
<50			×
≥50			×
Percent biopsy positivity (%)			
<34			×
≥34			×
No. of positive cores		×	
No. of negative cores		×	

CAPRA = Cancer of the Prostate Risk Assessment; MSKCC = Memorial Sloan Kettering Cancer; MSKCC_cores = Memorial Sloan Kettering Cancer Center nomogram with the number of positive and negative cores as the additional predictors in the model; PSA = prostate-specific antigen; RCS = restricted cubic splines.

2.3.2. Head-to-head comparison

To compare the prognostic performance of the different risk stratification tools, we used a split-sample approach. We first assigned each study participant to the appropriate risk category [2], [4], [6], [9], [10], [11], [12], calculated the CAPRA score [7], and computed the linear predictor for the preoperative MSKCC nomogram for BCR-free survival [13]. Then, we randomly split each imputed PCBaSe dataset into an equally sized training and testing dataset.

We restricted the main analysis to men with cT1c-cT3a tumors. For the D'Amico and EAU systems, which do not classify men with >cT2c tumors, men with cT3a tumors were classified to be at high risk in the main analysis. We also performed a sensitivity analysis restricted to men with cT1c-cT2c tumors. Furthermore, to evaluate the performance of the tools that also classify men with >cT3a tumors, we performed a second sensitivity analysis including men with cT1c-cT4 tumors.

We accounted for the presence of competing events by developing two separate cause-specific models in the training dataset: one for prostate cancer death and one for death from other causes. Cox proportional hazards models were used to estimate the cause-specific hazard ratios (HRs) and 95% confidence intervals (CIs) for prostate cancer death and death from other causes in the training datasets. The risk groups, CAPRA score, and MSKCC linear predictor were used as a single covariate in the models predicting prostate cancer death. The models predicting death from other causes included each risk stratification tool separately, age and year of diagnosis, Charlson

Comorbidity Index, marital status, education level, and primary treatment. Time at risk was calculated from the date of diagnosis until the date of death, emigration, or end of follow-up (December 31, 2016), whichever came first. Cause-specific hazards for prostate cancer death and death from other causes were combined to obtain cumulative incidence functions (CIFs) for prostate cancer death [22]. The estimated coefficients and CIFs were combined across the 50 imputed datasets [23].

Model performance was internally validated by computing discrimination and calibration in the testing datasets. Discrimination was evaluated by concordance index (C-index) adapted for competing risks [24], as described by Newson [25], in the full cohort and stratified by primary treatment (deferred treatment, curative treatment, and ADT) and by year of diagnosis (1998–2002, 2003–2006, and 2007–2016). The C-index was estimated by truncating the maximum follow-up time in the testing datasets at 1–19 yr of follow-up. Calibration was evaluated by comparing nonparametric CIFs [22] with the mean predicted CIFs within each category of the risk groups, CAPRA score, and each decile of the MSKCC linear predictor, at 5, 10, and 15 yr of follow-up.

As a sensitivity analysis for the multiple imputation, we performed the above-described analyses, except for the stratified analyses, for individuals diagnosed 2007 onward using a complete-case approach. We chose 2007 as the cutoff as NPCR started recording information on prostate volume and diagnostic biopsies in 2007. Furthermore, we evaluated the performance of the models predicting death from prostate cancer when competing risk was not taken into account.

All analyses were conducted in Stata (version 12.1; StataCorp, College Station, TX, USA).

3. Results

Baseline characteristics for 139 515 men included in the main analysis are presented in Table 3. During follow-up, 15 961 (11.44%) men died from prostate cancer.

Table 3. Baseline characteristics of men from the Prostate Cancer data Base Sweden (PCBaSe) 4.0 who were included in the main analysis.

	PCBaSe 4.0 (n = 139 515)	
	N	*%*
Age at diagnosis (yr), median (IQR)	69 (63–76)	
Year of diagnosis		
1998–2002	26 747	19.17
2003–2006	31 129	22.31
2007–2016	81 639	58.52
Mode of detection		
Health checkup	54 939	43.11
Lower urinary tract symptoms	39 270	30.81
Other symptoms	33 235	26.08
Missing	12 071	–
PSA (ng/ml), median (IQR)[a]	9.4 (5.8–20)	
Missing (*n*)	2110	
Prostate volume (ml), median (IQR)[b]	38 (29–52)	

	PCBaSe 4.0 ($n = 139\,515$)	
	N	%
Missing (n)	66 846	
Clinical tumor stage		
T1	65 804	49.37
T1a	5426	7.27
T1b	3493	4.68
T1c	65 682	88.04
Missing	122	–
T2	48 444	35.61
T3a	21 796	16.02
Missing[c]	3471	–
Biopsy Gleason score		
≤6	60 546	47.08
7	47 215	3671
3 + 4	28 680	65.95
4 + 3	14 810	34.05
Missing	3725	–
8	11 559	8.99
9	8552	6.65
10	729	0.57
Missing	10 914	–
Primary Gleason grade		
1	112	0.10
2	3706	3.17
3	80 229	68.62
4	30 237	25.86
5	2629	2.25
Missing	22 602	–
Secondary Gleason grade		
1	31	0.03
2	3517	3.01
3	65 608	56.20
4	39 704	34.01
5	7879	6.75
Missing	22 776	–
Number of cores sampled at biopsy, median (IQR)	10 (8–12)	
Missing (n)	44 118	
Total length of biopsy cores (mm), median (IQR)[d]	146 (119–172)	
Missing (n)	83 258	
Number of cores with cancer, median (IQR)[e]	3 (2–5)	
Missing (n)	44 826	

	PCBaSe 4.0 ($n = 139\,515$)	
	N	*%*
Total length of cancer (mm), median (IQR)[f]	9.4 (3–26)	
Missing (*n*)	77 667	
Primary treatment		
Deferred[g]	40 122	29.63
Curative[h]	60 496	44.68
Androgen deprivation	34 394	25.40
Death before treatment decision	383	0.28
Missing	4120	–
Charlson Comorbidity Index		
0	108 756	77.95
1	17 612	12.62
2	7971	5.71
3+	5176	3.71
Follow-up time (yr), median (IQR)[i]	5.83 (2.67–9.78)	
Cause of death		
Alive	93 337	66.90
Death from prostate cancer	15 961	11.44
Death from other causes	30 217	21.66

IQR = interquartile range (25–75th percentile); PSA = prostate-specific antigen.

a

PSA values of 0 ($n = 33$) truncated to minimum recorded value (0.1); PSA values over 10 000 ($n = 5$) truncated to 10 000.

b

Prostate volume of 0 ($n = 6$) truncated to minimum recorded value (2); prostate volume >1000 ($n = 10$) truncated to 1000.

c

Clinical tumor stage T0 ($n = 683$) recoded as missing.

d

Total length of biopsy cores of 0 ($n = 4$) truncated to minimum recorded value (1); total length of biopsy cores over 1000 ($n = 3$) truncated to 1000.

e

The number of cores with cancer recorded as 0 ($n = 24$) was recoded to 1; the number of cores with cancer larger than the total number of cores taken ($n = 2$) was recoded to equal the total number of cores taken.

f

    Total length of cancer of 0 ($n = 13$) truncated to the minimum recorded value (0.1); total length of cancer larger than the total length of biopsy cores ($n = 52$) was recoded to equal the total length of biopsy cores.

g

    Deferred treatment includes active surveillance or watchful waiting.

h

    Curative treatment includes radical prostatectomy or radiation therapy.

i

    Median follow-up time and IQR are reported for patients who did not die from prostate cancer ($n = 123\ 554$).

Across the compared risk stratification tools, the C-index at 10 yr ranged from 0.73 (95% CI: 0.72–0.73) to 0.81 (95% CI: 0.80–0.81; Fig. 1). The C-index generally increased with the granularity of the risk stratification tool, with the lowest discrimination for the three-tiered D'Amico, EAU, and NICE systems and the highest discrimination for the MSKCC nomograms, followed by the CAPRA score and the CPG system. The overall order of performance remained in analyses restricted to men with cT1c-cT2c tumors (Supplementary Fig. 1). In an analysis among men with cT1c-cT4 tumors, the MSKCC nomogram and the CPG and NCCN systems performed the best (Supplementary Fig. 2). Discrimination among men who received deferred and curative treatment was overall similar, except in the first 5 yr of follow-up (Fig. 2). Among curatively treated men, discrimination at 10 yr was better among men treated with radical prostatectomy (ranging from 0.74 [95% CI: 0.70–0.77] to 0.79 [95% CI: 0.76–0.83]) than among men treated with radiation therapy (ranging from 0.66 [95% CI: 0.63–0.68] to 0.73 [95% CI: 0.71–0.76]; Supplementary Fig. 3). Among men treated with primary ADT, discrimination was substantially poorer, ranging from 0.56 (95% CI: 0.55–0.56) to 0.65 (95% CI: 0.64–0.66). For all risk stratification tools, the discrimination improved in more recently diagnosed cohorts, ranging from 0.77 (95% CI: 0.76–0.78) to 0.86 (95% CI: 0.85–0.87) among patients diagnosed during 2007–2016 compared with 0.66 (95% CI: 0.65–0.67) to 0.74 (95% CI: 0.73–0.75) among patients diagnosed before 2003 (Supplementary Fig. 4).

Fig. 1. Pooled concordance index for prostate cancer death estimated in the testing datasets. The C-index was estimated by truncating follow-up time in testing dataset at 1–19 yr and plotted with the truncation year on the x axis.

AUA = American Urological Association; CAPRA = Cancer of the Prostate Risk Assessment; CPG = Cambridge Prognostic Groups; EAU = European Association of Urology; GUROC = Genito-Urinary Radiation Oncologists of Canada; MSKCC = Memorial Sloan Kettering Cancer; NCCN = National Comprehensive Cancer Network; NICE = The National Institute for Health and Care Excellence.

Fig. 2. Pooled concordance index for prostate cancer death estimated in the testing datasets, stratified by type of primary treatment. The C-index was estimated by truncating follow-up time in testing dataset at 1–19 yr and plotted with the truncation year on the x axis.

AUA = American Urological Association; CAPRA = Cancer of the Prostate Risk Assessment; CPG = Cambridge Prognostic Groups; EAU = European Association of Urology; GUROC = Genito-Urinary Radiation Oncologists of Canada; MSKCC = Memorial Sloan Kettering Cancer; NCCN = National Comprehensive Cancer Network; NICE = The National Institute for Health and Care Excellence.

Pooled coefficients, HRs, and corresponding 95% CIs for prostate cancer death in the training datasets for each risk stratification tool are reported in Supplementary Table 8. The observed and predicted CIFs in the testing dataset were close (Supplementary Table 9 and Supplementary Fig. 5–7), although the predicted CIFs seemed to be generally underestimated, especially at 5 yr of follow-up in the highest-risk category of the NCCN system, CAPRA score, and deciles of the MSKCC linear predictor.

In the sensitivity analysis for multiple imputation, the overall order of performance of the compared tools was the same in the complete-case and multiple imputation approaches; however, there were differences in the point estimates for most of the tools (Supplementary Tables 10 and 11, and Supplementary Fig. 8).

When competing risks were not taken into account, the C-indices were generally higher (Supplementary Fig. 9). The overall order of performance of the compared tools, however, remained. The observed and predicted CIFs were overall close, except for the highest risk category of the NCCN and CPG risk groups, CAPRA score, and deciles of the MSKCC linear predictor where the predicted CIFs were overestimated (Supplementary Table 12 and Supplementary Fig. 10–12).

4. Discussion

We systematically compared the prognostic performance of the most commonly used pretreatment risk stratification tools for prostate cancer. All tools showed rather good discrimination for prostate cancer death, with C-indexes at 10 yr of follow-up ranging from 0.73 to 0.81. In general, tools with more detailed risk stratification showed better discrimination. The MSKCC nomogram performed the best (C-index: 0.81), followed by the CAPRA score (C-index: 0.80) and the CPG system (C-index: 0.78).

We compared three types of risk stratification tools: risk grouping systems (D'Amico, EAU, NICE, GUROC, AUA, NCCN, and CPG), risk scores (CAPRA), and nomograms (MSKCC). Overall, the D'Amico and D'Amico-derived systems (ie, EAU, NICE, GUROC, AUA, and NCCN) performed similarly (C-index: 0.73–0.77). We observed higher discrimination in our study than previously reported for both the D'Amico (C-index: 0.73 vs 0.70) and the NICE (C-index: 0.73 vs 0.69 for internal and 0.66 for external validation) system [6], [26]. The EAU, GUROC, AUA, and NCCN systems have, to the best of our knowledge, never before been evaluated for predicting prostate cancer death. The CPG system, which was developed to predict prostate cancer death accounting for competing events [6], outperformed the other systems, with similar discrimination in our study to that in the original study (0.78 vs 0.75 for the internal and 0.79 for the external validation) and in a previous validation study [27].

Both the CAPRA score and the MSKCC nomogram have previously been validated for prostate cancer-specific death [28], [29], [30]. In our cohort, the C-index for the CAPRA score was 0.80, the same as in the two previous validation cohorts [28], [30]. It is important to note that although the MSKCC nomogram available on the MSKCC webpage predicts several outcomes, including prostate cancer death, the linear predictor and the baseline survival function are available for

predicting BCR only [13]. In our study, the MSKCC linear predictor was the best discriminating tool with a slightly lower C-index than in the validation study (C-index: 0.80 vs 0.82) [29].

Several of our study results are a consequence of well-established statistical principles, such as, in general, improved discrimination with more detailed risk stratification. However, our data also suggest that discrimination does not necessarily improve by simply subdividing the standard risk groups. For example, subdivision of the NCCN low-risk group into very low and low (C-index at 10 yr: 0.76 vs 0.76), intermediate-risk group into favorable and unfavorable (C-index at 10 yr: 0.76 vs 0.77), and high-risk group into very high and high (C-index at 10 yr: 0.75 vs 0.76) risk groups improved discrimination only slightly. Of note, the NCCN very-high-risk group identified men at a higher risk of dying from prostate cancer than the highest CPG or CAPRA score group.

Most tools used clinically and in guidelines are refined versions of the D'Amico system. Our data show that the D'Amico and D'Amico-derived tools are inferior to the MSKCC, CAPRA, and CPG tools. For example, using PSA, clinical tumor stage, and primary and secondary Gleason grades to predict individual probabilities of prostate cancer death using the MSKCC nomogram, compared with using the same information to categorize men into the three D'Amico risk groups, was, on its own, sufficient to improve discrimination from 0.73 to 0.80. This is an expected consequence of categorizing continuous data. In fact, using the whole range of PSA or Gleason score alone discriminates prostate cancer death better than the D'Amico system (C-index at 10 yr: 0.75 or 0.74 vs 0.73; Supplementary Fig. 13). Another way to illustrate this is the following: within the D'Amico intermediate-risk group, individual probabilities of prostate cancer death predicted using the MSKCC nomogram ranged from 1.9% to 41.1% at 15 yr of follow-up (Supplementary Fig. 14). Although fine risk stratification is not relevant for all prostate cancer patients, it may be highly relevant for some. For example, whether to add adjuvant ADT to radiotherapy can be a difficult treatment decision given the side effects of ADT. Individual risk prediction, coupled with data on the relative treatment benefits of adjuvant ADT from randomized trials, allows for estimation of absolute treatment benefits and better informed treatment decisions.

There are currently a plethora of available pretreatment risk stratification tools for prostate cancer. The prostate cancer community could gain by agreeing on adopting one or a few of the top performing tools. From a clinical practice perspective, more detailed risk prediction will allow for more personalized treatment decisions. From a research perspective, agreeing on using a specific tool as a benchmark or gold standard would improve comparability across studies. This will become even more important in the future due to developments in molecular pathology, imaging, and image-guided biopsy procedures.

The main strength of this study is the use of a nationwide population-based cohort encompassing almost 140 000 men with prostate cancer undergoing different primary treatments, with detailed clinicopathological data, long follow-up, and almost 16 000 recorded prostate cancer deaths. The downside of these real-world data is incomplete information on the predictors: although 80% of the study participants had complete information on PSA, Gleason grade, and clinical stage, only 35% had complete information on all variables used in some of the assessed risk stratification tools. Furthermore, information on cT2-cT3 substages is not recorded in the PCBaSe. We addressed the missing data problem by multiple imputation and by incorporating data from an external cohort of men with known cT substage. Misclassification is another limitation, especially among older men. For example, a proportion of the men with N0/NX status included in our study likely had true N1 disease, and this proportion is presumably higher among older men. However, such misclassification should not influence the overall order of performance of the compared risk stratification tools. It should also be noted that most men in this study were diagnosed in more recent years, and thus longer-term (eg, >15 yr) performance estimates are less precise. Moreover,

we were unable to formally externally validate the performance of the original models for the different tools in our data, as information on the intercept and/or the linear predictor have not been published, or because the tools were developed or validated to predict BCR and not prostate cancer death. It is possible that our internally validated model performance is overoptimistic. Our results may also not be generalizable to populations with different case mix without formal external validation.

5. Conclusions

In this study, we observed substantial differences in the prognostic performance of the most commonly used pretreatment risk stratification tools for prostate cancer. The MSKCC nomograms, followed by the CAPRA score and the CPG risk grouping system, performed better than the standard D'Amico-derived tools and are easy to apply in clinical practice. The use of these tools may improve clinical decision making.

References

[1]

M. Popiolek, J.R. Rider, O. Andren, *et al.***Natural history of early, localized prostate cancer: a final report from three decades of follow-up**

Eur Urol, 63 (2013), pp. 428-435

[2]

A.V. D'Amico, R. Whittington, S.B. Malkowicz, *et al.***Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer**

JAMA, 280 (1998), pp. 969-974

[3]

J.I. Epstein, P.C. Walsh, M. Carmichael, C.B. Brendler**Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer**

JAMA, 271 (1994), pp. 368-374

[4]

J.L. Mohler, A.J. Armstrong, R.R. Bahnson, *et al.***Prostate cancer, version 1.2016**

J Natl Compr Canc Netw, 14 (2016), pp. 19-30

[5]

Z.S. Zumsteg, Z. Chen, L.E. Howard, *et al.***Number of unfavorable intermediate-risk factors predicts pathologic upstaging and prostate cancer-specific mortality following radical prostatectomy: results from the SEARCH database**

Prostate, 77 (2017), pp. 154-163

[6]

V.J. Gnanapragasam, A. Lophatananon, K.A. Wright, K.R. Muir, A. Gavin, D.C. Greenberg**Improving clinical risk stratification at diagnosis in primary prostate cancer: a prognostic modelling study**

PLoS Med, 13 (2016), Article e1002063

[7]

M.R. Cooperberg, D.J. Pasta, E.P. Elkin, *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, 173 (2005), pp. 1938-1942

[8]

A.J. Stephenson, P.T. Scardino, J.A. Eastham, *et al.***Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy**

J Natl Cancer Inst, 98 (2006), pp. 715-717

[9]

J. Graham, P. Kirkbride, K. Cann, E. Hasler, M. Prettyjohns**Prostate cancer: summary of updated NICE guidance**

BMJ, 348 (2014), p. f7524

[10]

H. Lukka, P. Warde, T. Pickles, *et al.***Controversies in prostate cancer radiotherapy: consensus development**

Can J Urol, 8 (2001), pp. 1314-1322

[11]

M.G. Sanda, J.A. Cadeddu, E. Kirkby, *et al.***Clinically localized prostate cancer: AUA/ASTRO/SUO guideline. Part I: risk stratification, shared decision making, and care options**

J Urol, 199 (2018), pp. 683-690

[12]

N. Mottet, J. Bellmunt, M. Bolla, *et al.***EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent**

Eur Urol, 71 (2017), pp. 618-629

[13]

Memorial Sloan Kettering Cancer Center. Dynamic prostate cancer nomogram: coefficients. Memorial Sloan Kettering Cancer Center site.

[14]

E. Hagel, H. Garmo, A. Bill-Axelson, *et al.***PCBaSe Sweden: a register-based resource for prostate cancer research**

Scand J Urol Nephrol, 43 (2009), pp. 342-349

[15]

M. Van Hemelrijck, A. Wigertz, F. Sandin, *et al.***Cohort profile: the National Prostate Cancer Register of Sweden and Prostate Cancer data Base Sweden 2.0**

Int J Epidemiol, 42 (2013), pp. 956-967

[16]

K. Tomic, A. Berglund, D. Robinson, *et al.***Capture rate and representativity of the national Prostate Cancer Register of Sweden**

Acta Oncol, 54 (2015), pp. 158-163

[17]

J.F. Ludvigsson, C. Almqvist, A.K. Bonamy, *et al.***Registers of the Swedish total population and their use in medical research**

Eur J Epidemiol, 31 (2016), pp. 125-136

[18]

K. Fall, F. Stromberg, J. Rosell, O. Andren, E. Varenhorst**South-East Region Prostate Cancer group. Reliability of death certificates in prostate cancer patients**

Scand J Urol Nephrol, 42 (2008), pp. 352-357

[19]

A. Marshall, D.G. Altman, R.L. Holder**Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study**

BMC Med Res Methodol, 10 (2010), p. 112

[20]

I.R. White, P. Royston, A.M. Wood**Multiple imputation using chained equations: issues and guidance for practice**

Stat Med, 30 (2011), pp. 377-399

[21]

S. Johansson, L. Astrom, F. Sandin, U. Isacsson, A. Montelius, I. Turesson**Hypofractionated proton boost combined with external beam radiotherapy for treatment of localized prostate cancer**

Prostate Cancer, 2012 (2012), Article 654861

[22]

V. Coviello, M. Boggess**Cumulative incidence estimation in the presence of competing risks**

Stata J, 4 (2004), pp. 103-112

[23]

A. Marshall, D.G. Altman, R.L. Holder, P. Royston**Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines**

BMC Med Res Methodol, 9 (2009), p. 57

[24]

M. Wolbers, M.T. Koller, J.C. Witteman, E.W. Steyerberg**Prognostic models with competing risks: methods and application to coronary risk prediction**

Epidemiology, 20 (2009), pp. 555-561

[25]

R.B. Newson**Comparing the predictive powers of survival models using Harrell's C or Somers' D**

Stata J, 10 (2010), pp. 339-358

[26]

S.A. Boorjian, R.J. Karnes, L.J. Rangel, E.J. Bergstralh, M.L. Blute**Mayo Clinic validation of the D'amico risk group classification for predicting survival following radical prostatectomy**

J Urol, 179 (2008), pp. 1354-1360

[27]

V.J. Gnanapragasam, O. Bratt, K. Muir, *et al.***The Cambridge Prognostic Groups for improved prediction of disease mortality at diagnosis in primary non-metastatic prostate cancer: a validation study**

BMC Med, 16 (2018), p. 31

[28]

M.R. Cooperberg, J.M. Broering, P.R. Carroll**Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis**

J Natl Cancer Inst, 101 (2009), pp. 878-887

[29]

A.J. Stephenson, M.W. Kattan, J.A. Eastham, *et al.***Prostate cancer-specific mortality after radical prostatectomy for patients treated in the prostate-specific antigen era**

J Clin Oncol, 27 (2009), pp. 4300-4305

[30]

J.M. Vainshtein, M. Schipper, S. Vance, F.Y. Feng, K.B. Olson, D.A. Hamstra**Limitations of the Cancer of the Prostate Risk Assessment (CAPRA) prognostic tool for prediction of metastases and prostate cancer-specific mortality in patients treated with external beam radiation therapy**

Am J Clin Oncol, 39 (2016), pp. 173-180