

Meta-analysis of predictive models to assess the clinical validity and utility for patient-centered medical decision making: application to the CAncer of the Prostate Risk Assessment (CAPRA)

Marine Lorent, Haïfa Maalmi, Philippe Tessier, Stéphane Supiot, Etienne Dantan, Yohann Foucher

▶ **To cite this version:**

Marine Lorent, Haïfa Maalmi, Philippe Tessier, Stéphane Supiot, Etienne Dantan, et al.. Meta-analysis of predictive models to assess the clinical validity and utility for patient-centered medical decision making: application to the CAncer of the Prostate Risk Assessment (CAPRA). BMC Medical Informatics and Decision Making, 2019, 19 (1), pp.2. 10.1186/s12911-018-0727-2 . inserm-02306977

HAL Id: inserm-02306977

https://inserm.hal.science/inserm-02306977

Submitted on 7 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Meta-analysis of predictive models to assess the clinical validity and utility for patient-centered medical decision making: application to the CAncer of the Prostate Risk Assessment (CAPRA)

Marine Lorent[1], Haïfa Maalmi[2], Philippe Tessier[1], Stéphane Supiot[3,4], Etienne Dantan[1] and Yohann Foucher[1,5,6*]

Abstract

Background: The Cancer of the Prostate Risk Assessment (CAPRA) score was designed and validated several times to predict the biochemical recurrence-free survival after a radical prostatectomy. Our objectives were, first, to study the clinical validity of the CAPRA score, and, second, to assess its clinical utility for stratified medicine from an original patient-centered approach.

Methods: We proposed a meta-analysis based on a literature search using MEDLINE. Observed and predicted biochemical-recurrence-free survivals were compared to assess the calibration of the CAPRA score. Discriminative capacities were evaluated by estimating the summary time-dependent ROC curve. The clinical utility of the CAPRA score was evaluated according to the following stratified decisions: active monitoring for low-risk patients, prostatectomy for intermediate-risk patients, or radio-hormonal therapy for high risk patients. For this purpose, we assessed CAPRA's clinical utility in terms of its ability to maximize time-dependent utility functions (i.e. Quality-Adjusted Life-Years – QALYs).

Results: We identified 683 manuscripts and finally retained 9 studies. We reported good discriminative capacities with an area under the SROCt curve at 0.73 [95%CI from 0.67 to 0.79], while graphical calibration seemed acceptable. Nevertheless, we also described that the CAPRA score was unable to discriminate between the three medical alternatives, i.e. it did not allow an increase in the number of life years in perfect health (QALYs) of patients with prostate cancer.

Conclusions: We confirmed the prognostic capacities of the CAPRA score. In contrast, we were not able to demonstrate its clinical usefulness for stratified medicine from a patient-centered perspective. Our results also highlighted the confusion between clinical validity and utility. This distinction should be better considered in order to develop predictive tools useful in practice.

Keywords: Meta-analysis, Prostate cancer, Patient-centered outcomes, Stratified medicine

* Correspondence: Yohann.Foucher@univ-nantes.fr
[1]SPHERE (methodS in Patient-centered outcomes & HEalth ResEarch) U1246, INSERM, Nantes University, Tours University, Nantes, France
[5]Nantes University Hospital, Nantes, France
Full list of author information is available at the end of the article

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

Prostate cancer (PC) is the most common cancer in men, accounting for approximately 10% of all male cancer deaths in economically developed countries [1]. In the long term, many patients experience biochemical recurrence (BCR), defined by an increase in prostate-specific antigen (PSA) levels [2], which can also lead to cancer-specific mortality [3]. Many developments have been made in predictive scoring systems [4–7] in order to avoid both over-treatment of patients at low-risk of cancer recurrence or death related to the cancer, and the under-treatment of patients at high-risk.

Among these scores, the Cancer of the Prostate Risk Assessment (CAPRA) score [8] was designed to predict the pre-operative risk of BCR-free survival after a radical prostatectomy (RP). RP represents the most widely performed curative therapy for patients with localized PC [9]. The CAPRA score is calculated from five variables measured before prostatectomy: the PSA level, the Gleason score, the clinical T stage, the percentage of positive prostate biopsies and the patient age at diagnosis. Meurs et al. [10] performed a substantial meta-analysis of seven studies and concluded that the CAPRA rule accurately predicts BCR-free survival at 3 years.

Nevertheless, despite its good predictive capacities, the clinical usefulness of the CAPRA score may be questioned. Firstly, the results of the meta-analysis of Meurs et al. [10] mainly consisted in risk ratios between expected and observed numbers of events among three CAPRA-based strata (0–2 for low risk, 3–5 for intermediate risk, and 6–10 for high risk). The aim was to evaluate the calibration compared to the initial study of Cooperberg et al. [8]. But the discriminative capacities were not evaluated. Secondly, the CAPRA thresholds used to define the three strata were arbitrarily defined, while the purpose of such a scoring system is to adapt the medical management of patients regarding the strata they belongs to. In two available online applications related to the CAPRA computation [11, 12], the following recommendations are proposed: routine surveillance for low risk patients, localized treatment (surgery or radiation alone, brachytherapy with or without external-beam therapy) for intermediate-risk patients, hormonal therapy or multimodal therapy (surgery with radiation, or radiation therapy with hormonal therapy) for high-risk patients. However, to the best of our knowledge, no study has assessed the clinical usefulness of the CAPRA score for stratified medical decision making.

We hypothesized that the clinical usefulness depends on its ability to bring benefits for patients by differentiating the following alternative therapies: the combined radio-hormonal therapy (RHT) for patients at high-risk of BCR, the RP for medium-risk patients and the active monitoring (AM) for low-risk patients. In this paper, from an update of the meta-analysis proposed by Meurs et al. [10], our objectives were i) to study the *clinical validity* of the CAPRA score, i.e. its prognostic capacities both in terms of discrimination and calibration, and ii) to evaluate its *clinical utility*, i.e. its usefulness for stratified medical decision making from an original patient-centered methodology that considers its corresponding benefits and drawbacks [13].

Methods

Search strategy

We conducted a literature search using the MEDLINE citation database to identify eligible studies by combining the following MeSH terms and keywords: “CAPRA score”, “biochemical recurrence” and “prostate cancer”. Since our analysis was an update of the meta-analysis of Meurs et al. [10] published in March 2012, we restricted our search after this date. No language restriction was performed. Cross-referencing was applied to complete the identification of other studies.

Study selection

Two investigators (Y.F and E.D) independently screened the titles and abstracts of papers identified from the literature search. Irrelevant articles were excluded. The full text of each potentially relevant article was reviewed and checked independently by two persons (Y.F and E.D or H.M). Discrepancies between the three investigators were resolved by consensus. As illustrated in Table 1, a study was included in the analysis if it met the two following criteria: 1) prospective or retrospective study conducted in PC patients having experienced RP, with the preoperative CAPRA score as a predictor and BCR-free survival as the endpoint; 2) study delivering Hazard Ratios (HRs) and/or BCR-free survival curves stratified according to CAPRA values. When the same cohort was used in several publications, only the most recent study was selected if it contained the statistical indicators necessary for our meta-analysis.

Data extraction

From the eligible studies, two independent investigators (H.M and Y.F) collected the data independently. For each study, the following characteristics were collected:

Table 1 PICOS table related to the selection of the papers

Patient population	Patients with Prostate Cancer with available CAPRA score
Intervention	Having a radical prostatectomy
Comparison intervention	With or without control group
Outcomes	Biochemical-recurrence-free survival
Study type	Prospective or retrospective study presenting hazard ratio and/or survival curves

first author, year of publication, country, period of recruitment, duration of follow-up, study setting and design, sample size, age, reported outcomes, association between CAPRA and BCR-free survival (HRs and 95%CI), number of patients in each CAPRA-based strata, presence of survival curves and the corresponding number of at-risk patients, and the Harrell's concordance index (C-index). Any disagreement was resolved by consensus. All statistical analyses were performed using the 3.1.1 version of the R software [14]. The survival probabilities were extracted from a digitalized picture by using the R packages *ReadImages* and *digitize* [15]. Many papers do not provide the number of at-risk patients over time, so we applied the method proposed by Parmar et al. [16] in order to obtain estimations.

Assessing clinical validity

The association between the CAPRA score and BCR-free survival was evaluated by estimating pooled HRs which compared the high-risk category (CAPRA ≥6) and the intermediate-risk category (2 < CAPRA < 6) to the low-risk category of the CAPRA score (CAPRA≤2). When it was available, we used the category from 0 to 2 as the low-risk category, or otherwise we used the category 0 to 1. Collected HRs were pooled by applying a random-effects model and the DerSimonian-Laird estimator [17] (R package *metaphor*). Sensitivity analysis was conducted by eliminating a single study one-at-a-time. Publication bias was evaluated statistically by the Kendall's tau and the Egger's linear regression test of the intercept [18]. Heterogeneity between studies was assessed by Cochrane Q and I^2 statistics [19] (R package *meta*). We also investigated potential heterogeneity by comparing pooled HRs stratified given possible heterogeneity factors: effective size, inclusion period and the geographic origin of the study. The outcome definition appeared homogeneous between studies, so there was no reason to test it as a heterogeneity factor.

Pooled survival curves per CAPRA-based strata were estimated by the distribution-free approach with random effects proposed by Combescure et al. [20]. The calibration was assessed graphically by comparing the pooled survival curves for each CAPRA strata to those estimated in the original paper by Cooperberg [8]. The discrimination capacities at 5-years post-RP were evaluated using the time-dependent summary ROC (SROCt) curves [21].

Assessing clinical utility

We endorse a patient-centered approach to medical decision making by defining the clinical utility of a decision in terms of the expected Quality-Adjusted Life-Years (QALYs) it brings. QALYs combine into a single index, information about the length of life and health-related quality of life. They are computed by weighting each period of life (usually a year) by a utility score such that a score of zero indicates death and a score of one stands for perfect health [22]. The utility scores represent the individuals' preferences over health states such that a higher score indicates a more preferred health state. Although QALYs were primarily designed for the conduct of economic evaluation, their composite nature may also be useful to inform clinical decision making [23].

More precisely, as demonstrated by Dantan et al. [24], the expected number of QALYs resulting from the use of a prognostic marker's threshold to decide between two alternative therapies can be determined as follows. Let T, X and κ be the time-to-failure, the baseline prognostic marker under investigation to drive the treatment allocation, and a given threshold of X. Let $D(\tau)$ be the indicator of failure such that $D(\tau) = 1$ if it occurs before the forecast horizon time τ, and $D(\tau) = 0$ otherwise. Patient profiles may therefore be distinguished by combining their two possible initial strata according the marker X ($X > \kappa$ or $X \leq \kappa$) and the two possible outcomes ($D(\tau) = 1$ or $D(\tau) = 0$) over the time horizon under consideration. The optimal threshold κ^* can be estimated as the threshold that maximizes the expected number of QALYs:

$$\sum\nolimits_{g\in\{X>\kappa, X\leq\kappa\}} P(g)\left[u_{g,0}E(\min(T,\tau)|g) + u_{g,1}(\tau - E(\min(T,\tau)|g))\right] \quad (1)$$

where $P(g)$ is the proportion of patients in strata g, $E(\min(T, \tau)|\, g)$ is the Restricted Mean Survival Time (RMST) up to time τ in the strata g, i.e. the average survival time when patients are followed up to τ [25] and $(u_{g,\,0}, u_{g,\,1})$ are the utility scores for health states $D(\tau) = 0$ and $D(\tau) = 1$ in the strata g. Therefore, calculating QALYs requires information about i) the survival probabilities and ii) the utility scores corresponding to the relevant health states.

Survival estimation

The RMST, $E(\min(T, \tau)|\, g)$, which need to be estimated for all possible thresholds κ of the marker X, was estimated as the area under the corresponding survival curve $S(t|\, g)$ until time τ. We used the methodology previously proposed by Combescure et al. [21]. Survival probability at time t given strata g is defined as follows and numerically calculated using the 30-points Gauss–Legendre quadrature:

$$S(t|g = \{X > \kappa\}) = \frac{\int_\kappa^\infty \exp\left(-\int_0^t E_v[\lambda(u|x,v)]du\right)E[f(x|\omega)]dx}{\int_\kappa^\infty E[f(x|\omega)]dx} \quad (2)$$

In eq. 2, the distribution of the marker X, $f(x|\, \omega_i)$, is estimated according to a log-normal distribution interval-truncated in-between 0 and 10 (i.e. the CAPRA range),

$\omega_i \sim N(0, \sigma_\omega)$ is a random effect of the mean for the study i and σ_ω its standard error. The instantaneous risk function of BCR, namely $\lambda(t|\, x, \omega_i)$, is estimated according to a flexible piecewise constant risk function:

$$\lambda(t|x, v_i) = exp\left[v_i + \sum_{l=0}^{L}\left(\beta_{l,1} + \beta_{l,2}\right) \times 1\{t > \tau_l\}\right] \quad (3)$$

where $(\beta_{l,\,1}, \beta_{l,\,2})$ are the regression coefficients specific to the interval $l \in \{1, \ldots, L\}$ (L being the number of time intervals), $v_i \sim N(0, \sigma_v)$ is a random effect of the mean for the study i and σ_v its standard error. To obtain $f(x|\, \omega_i)$ and $\lambda(t|\, x, v_i)$, we used the R package *nlme* based on the maximization of the restricted log-likelihood.

The survival probabilities were estimated from articles including patients with RP. For a stratified decision, we considered alternative therapies (AM and RHT) that will impact the survival probabilities. We studied several plausible scenarios in terms of decreasing length of life: a RMST loss of 10% at 5 years for AM versus RP [26] and in-between 20 to 40% for RP versus RHT [27–29].

Utility scores determination

Following Koerber et al. [30], we assumed that the baseline utility for PC patients can be approximated by an age-adjusted utility score of the general population. This baseline utility is then combined with the utilities of various health conditions (symptoms of prostate cancer and treatments) using a multiplicative model. Although there is no consensus about the appropriate method that should be used, there is recent evidence in favor of the multiplicative model for combining scores for health states compared to the alternative model [31, 32].

The event predicted by the CAPRA score is the first failure between disease progression (biochemical recurrence or distant metastasis) and death. For each treatment group (RP, AM, RHT), we calculated the expected utility before this event as the product of age-adjusted baseline utility and the utilities of symptoms associated with the treatment group weighted by their repartition. We calculated the corresponding utility after the event as the product of utility before the event and the utilities of disease progression or death weighted by their repartition. Supplementary information on the precise method of utility calculation is provided in the Additional file 1.

Numerous studies have published utility scores for health states related to PC assessed using various methods [33–38]. We chose the utility scores estimated by Stewart et al. [37] on a sample of patients aged 60 (the mean/median patient age in the included study in our meta-analysis centered around this value, Table 3) and over. These scores were deemed the most suitable for our study because they were directly assessed from patients and they were reported for both single symptoms and for comorbid health conditions. The raw utility scores and their sources are reported in Table 2. Since there was no utility value for active monitoring in the study proposed by Stewart et al. [37], we arbitrarily chose a score of 0.98 which seems in line with recent evidence suggesting that men undergoing active surveillance have high utility scores [33]. Repartition of each symptom and subtype of event given the treatment group are based on a literature search and are detailed in the Additional file 1.

Results

Description of the selected studies

A flowchart of the selected studies is presented in Fig. 1. We identified 683 manuscripts of which only 12 underwent full text review. No additional study was identified by cross-referencing. We also considered the 8 validation studies included in the meta-analysis of Meurs et al. [10] Among these 20 articles, we finally retained 9 studies: the original study of Cooperberg et al. [8], 5 studies [39–43] published between 2005 and 2012 and selected by Meurs et al. [10], and 3 additional studies [44–46] published more recently. The study proposed by Tamblyn et al. [47] in 2011, selected by Meurs et al. [10], was excluded because the authors did not present BCR-free survival curves according to CAPRA values. The work of Lughezzani et al. [48] in 2010, selected by Meurs et al. [10], was excluded because the same population with a longer follow-up was used by Budaus et al. [44] in 2012. Furthermore, the latter included all the statistical indicators necessary for our meta-analysis.

Table 2 Sources and estimation of raw utility scores[a]

	Value	Source
Baseline utility value for age 60	0.84	Ara and Brazier [32]
Sexual dysfunction	0.89	Stewart et al. [37]
Sexual dysfunction and urinary incontinence	0.78	
Impotence and bowel dysfunction	0.57	
Impotence, urinary incontinence, and bowel dysfunction	0.45	
Local disease progression	0.67	
Metastatic cancer	0.25	
Active monitoring	0.98	Authors' assumption

[a]The table reports 'raw' scores. The final scores used in our calculations were obtained by combining the baseline score with these raw scores using a multiplicative model. For instance, the final score for sexual dysfunction is obtained by multiplying the score in Stewart et al. by our baseline utility value (see the Additional file 1)
Finally, for patients alive without BCR, the utility scores were estimated at 0.79 under AM, 0.76 under RP and 0.72 under RHT. By merging the utilities related to the combined event (disease progression or death), the expected utility scores after the treatment failure were assessed at 0.34 under AM, 0.24 under RP and 0.20 under RHT. A sensitivity analysis was carried out by varying the patient age between 55 and 75 years. The conclusions of the analyses were unchanged (data not shown)

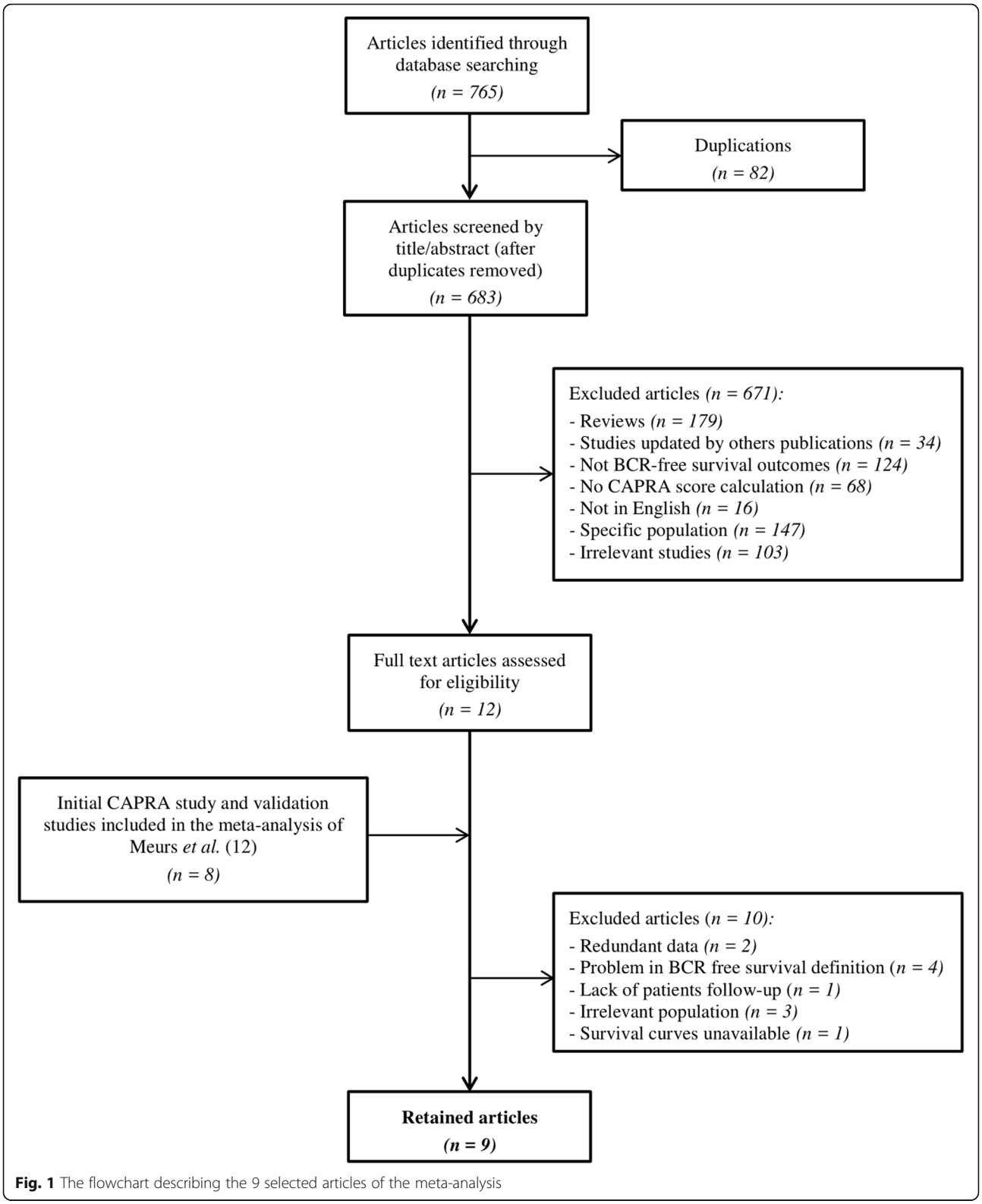


Fig. 1 The flowchart describing the 9 selected articles of the meta-analysis

The main features of the 9 eligible studies (15,908 patients) are summarized in Table 3. All studies were based on prospective cohorts of patients treated by RP. The sample size per study ranged from 115 to 6737 patients. The mean age ranged from 58.0 to 66.4 years. Four studies were conducted

Table 3 Description of the 9 studies included in the meta-analysis

Study	Country, Period, Follow-up	Effective size, Ethnicity, Age	Definition of BCR	CAPRA (N)	HR (95%CI)	C-index
Cooperberg et al., 2005 [8]	USA, 1992–2001, median = 24 m	1439, 88% Caucasian 8.4% African-American mean = 62y (no SD)	2 consecutive PSA ≥ 0.2 ng/mL or secondary treatment for elevated postoperative PSA	0–1 (401) 2 (432) 3 (296) 4 (155) 5 (84) 6 (43) 7–10 (28)	1.28 (0.79–2.08) 2.36 (1.49–3.72) 2.38 (1.40–4.03) 3.32 (1.89–5.80) 7.11 (3.84–13.15) 17.38 (9.92–30.46)	0.660
Cooperberg et al., 2006 [39]	USA, 1988–2004, mean = 42 m	1309, 59% Caucasian, mean = 61.9y (±6.6)	1 PSA > 0.2 ng/mL or 2 consecutive PSA ≥ 0.2 ng/mL or secondary treatment for elevated postoperative PSA	0–1 (324) 2 (329) 3 (291) 4 (158) 5 (108) 6 (76) 7–10 (60)	1.89 (1.25–2.85) 2.75 (1.85–4.10) 3.29 (2.15–5.04) 4.51 (2.89–7.05) 7.19 (4.58–11.30) 9.90 (6.34–15.46)	0.680
May et al., 2007 [42]	Germany, 1992–2005, mean = 56 m	1296, 100% Caucasian, mean = 63.7y (±5.5)	1 PSA > 0.2 ng/mL, 2 PSA ≥ 0.2 ng/mL or secondary treatment for elevated postoperative PSA	0–1 (130) 2 (297) 3 (265) 4 (222) 5 (140) 6 (92) 7–10 (150)	unknown	0.810
Zhao et al., 2008 [43]	USA, 1984–2006, median = 4y	6737, 91% Caucasian, mean = 58y (no SD)	1 PSA level > 0.2 ng/mL	0–1 (2796) 2 (1937) 3 (1005) 4 (463) 5 (306) 6 (161) 7–10 (69)	2.24 (1.83–2.74) 3.69 (3.01–4.54) 8.61 (6.96–10.65) 9.52 (7.57–11.97) 13.41 (10.40–17.29) 18.96 (13.87–25.92)	0.760
Loeb et al., 2012 [41]	USA, 2003–2009, median = 34 m	726, 93.5% Caucasian, mean = 59.3y (no SD)	Repeated PSA ≥ 0.2 ng/mL, secondary treatment for elevated postoperative PSA	0–1 (441) 2 (263) 3 (113) 4 (80) 5 (57) 6 (23) 7–10 (13)	1.1 (0.4–3.0) 4.1 (1.7–10.0) 5.1 (2.1–12.5) 10.5 (4.4–24.7) 12.2 (3.8–38.9) 47.8 (17.1–133.2)	0.764
Ishizaki et al., 2011 [40]	Japan, 1999–2010, mean = 38 m	211, 100% Japanese, mean = 62.2y (±5.8)	2 PSA ≥ 0.2 ng/mL, secondary treatment for elevated postoperative PSA	0–2 (85) 3–5 (106) 6–10 (20)	2.14 (1.19–3.86) 9.14 (4.30–19.44)	0.755
Budaus et al., 2012 [44]	Germany, 1992–2009, mean = 56 m	2937, unknown, median = 64y (48–74)	1 PSA level ≥ 0.2 ng/mL	0–2 (1280) 3–5 (1270) 6–10 (387)	3.1 (2.4–3.9) 7.0 (5.5–9.0)	0.714
Yoshida et al., 2012 [46]	Japan, 1995–2008, median = 44 m	503, 100% Japanese, median = 65y (47–76)	1 PSA ≥ 0.2 ng/mL followed by a 2nd PSA higher, radiotherapy or hormonal therapy for the postoperative PSA elevation	0–2 (138) 3–5 (257) 6–10 (108)	1.67 (0.93–2.99) 3.97 (2.18–7.24)	0.673
Seo et al., 2014 [45]	Korea, 2008–2013, median = 13 m	115, 100% Korean, mean = 66.4y (±6.5)	2 PSA ≥ 0.2 ng/mL, additional treatment more than 6 months after RP	unknown	unknown	0.770

BCR Biochemical Recurrence, *CAPRA* Cancer Prostate Risk Assessment, *95%CI* Confidence Interval, *C-index* Harrell's concordance index, *HR* Hazard Ratio, *m* months, *SD or* ± Standard Deviation, *PSA* Prostate-Specific Antigen, *RP* Radical Prostatectomy, *y* years

in the USA, 2 in Germany, 2 in Japan, 1 in Korea and 1 in Australia.

Association between the CAPRA score and BCR-free survival

The 9 retained articles [8, 39–46] provided HRs or survival curves (from which one can re-estimate the HRs) according to CAPRA-based strata. As illustrated in Fig. 2, the pooled HR of patients at high-risk (CAPRA≥6) versus those at low-risk (CAPRA≤2) was 10.65 [95%CI from 6.63 to 17.10]. The HR was 3.29 [95%CI from 2.18 to 4.96] for patients at intermediate-risk compared to those at low-risk. The analysis indicated a non-significant publication bias (Kendall's test statistic = 0.28, $p = 0.3585$ for the two group comparisons; Egger's tests = 1.06, $p = 0.210$ for high versus low risk groups; Egger's tests = 0.46, $p = 0.6474$ for intermediate versus low risk groups) and a large heterogeneity among the studies (Cochrane Q = 63.16, $p < 0.0001$ for high versus low risk groups; Cochrane Q = 33.08, $p < 0.0001$ for intermediate versus low risk groups; $I^2 = 91.18$ and 84.25% for the 2 respective group comparisons). As illustrated in Table 4, one cannot identify any significant reason to explain this heterogeneity.

Clinical validity of the CAPRA score up to five years post-prostatectomy

The area under the SROCt curve (AUC) at 5 years post-prostatectomy was 0.73 [95%CI from 0.67 to 0.79]. The AUCs for each study ranged from 0.67 [95%CI from 0.61 to 0.73] to 0.76 [95%CI from 0.71 to 0.81], indicating acceptable discriminative capacities of the CAPRA score.

Seven validation studies [39–41, 44–46] with survival curves according to the 3 CAPRA-based strata were considered. As illustrated in Fig. 3, patients at low-risk have 80.8% chance [95%CI from 70.8 to 86.6] of not suffering a BCR or a death within the 5 years post-RP. In contrast, patients at high-risk have a 73.2% risk [95%CI from 61.7 to 86.6] of suffering a BCR or death within 5 years post-RP. Note that in the original study, the 5-year BCR-free survival probabilities in low-, intermediate-, and high-risk strata were 84.7, 62.7 and 21.8%, respectively (Fig. 3, 95%CI were not provided). In comparison to the pooled survival curves, the calibration of the CAPRA score appeared acceptable.

Clinical utility of the CAPRA score for deciding AM instead of RP

Only patients with a CAPRA score strictly lower than 6 were included since AM is not proposed to high-risk patients (i.e. CAPRA≥6). The maximum number of QALYs was obtained when RP is proposed to all patients. In other words, for all the CAPRA thresholds in-between 1 and 5, the expected gain in terms of QoL related to AM was insufficient to balance the associated decrease in the BCR-free survival. More precisely, the mean number of QALYs, i.e. the expected equivalent of years lived in perfect health, was 42.4 months if all patients are treated by RP. In contrast, when patients with a CAPRA score lower than 2 (the usual threshold separating low and intermediate-risk strata) are treated by AM and the other ones by RP, the mean number of QALYs was 41.7 months. Therefore, for a cohort of 100 patients, this stratification was associated with an expected loss of 70 months in perfect health compared to treating all patients by RP.

Clinical utility of the CAPRA score for deciding RHT instead of RP

Only patients with a CAPRA score strictly higher than 2 were included since treatment by RHT is not proposed to low-risk patients (i.e. CAPRA≤2). The maximum number of QALYs was obtained by proposing RHT to all patients: 41.6 months in perfect health state. Regardless of the CAPRA threshold, all the stratified medical decision making resulted in lower numbers of QALYs. For instance, by using the usual CAPRA-based stratification (CAPRA< 6 for intermediate risk strata versus CAPRA≥6 for high risk strata), we estimated that the mean number of life-years in perfect health was 39.6 months. Therefore, for a cohort of 100 patients, the use of CAPRA threshold at 6 for stratifying the medical decision between RP and RHT would have been associated with a loss of 200 months in perfect health state compared to treating all patients by RHT.

Discussion

Prognostic tools are essential in PC to adapt the medical management regarding the risk-level of patients [49–51]. In this context, our meta-analysis aimed to precisely describe the usefulness of the CAPRA score for stratified medicine.

Our results were obtained by analyzing 9 studies. In agreement with the literature and by considering the three usual CAPRA-based strata, we validated the association between the CAPRA score and the BCR-free survival (HR intermediate versus low = 3.29 [95%CI from 2.18 to 4.96] and HR high versus low = 10.65 [95%CI from 6.63 to 17.10]). We additionally described the CAPRA score as a marker with a robust capacity to discriminate BCR or dead patients at 5-year post-RP from alive patients without BCR at this time (AUC = 0.73 [95%CI from 0.67 to 0.79]). In terms of calibration, we also reported acceptable properties up to 5 years post-RP. However, despite this validity, we were unable to demonstrate the clinical utility of the CAPRA score in regards to the following stratified medical decision making: AM for low-risk patients, RP for intermediate-risk patients and RHT for high-risk patients.

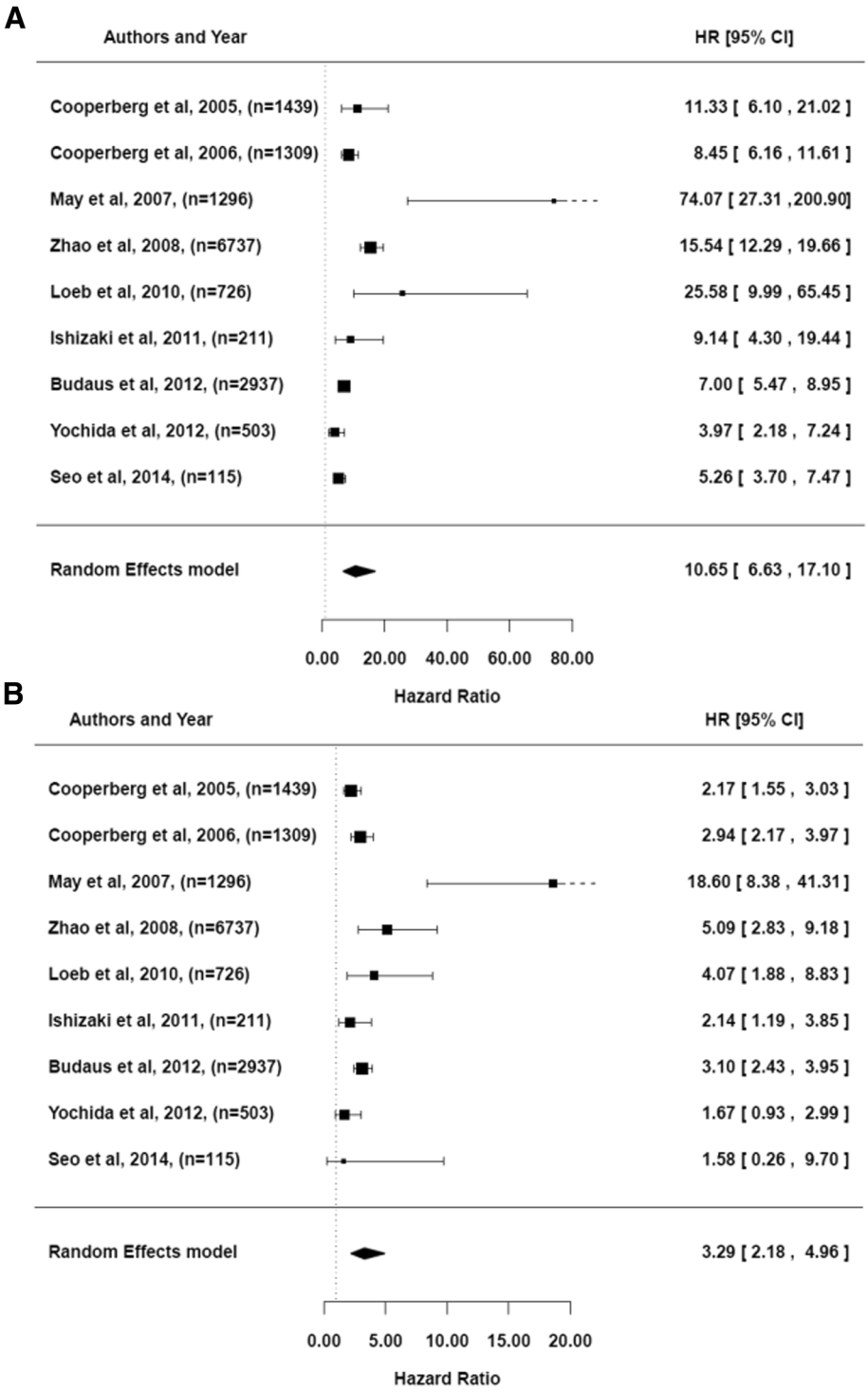


Fig. 2 The forest plots for pooled hazard ratios. **a**) High-risk group (CAPRA ≥6) versus the low-risk group (CAPRA ≤2). **b**) Intermediate-risk group (2 < CAPRA < 6) versus the low-risk group (CAPRA ≤2)

Our study highlights the important confusion between clinical validity and utility. Indeed, even if the relevance of stratified medical decision making partially depends on the prognostic capacities of a biomarker or a scoring system, other dimensions are also important to consider, in particular the consequences of both quantity and quality of life, and on trade-offs between them. Several recent studies discussed the importance of considering

Table 4 Investigation of the heterogeneity of the HRs according to CAPRA-based strata

	Number of studies per subgroup	Pooled HR (High versus Low) per subgroup	*p*-value	Pooled HR (Intermediate versus Low) per subgroup	*p*-value
Effective size	> 1000: 5 ≤ 1000: 4	14.60 [7.24–29.43] 8.13 [4.13–16.01]	0.240	4.45 [2.32–8.54] 2.35 [1.59–3.47]	0.098
Inclusion period	Inclusion start ≥1999: 3 Inclusion start < 1999: 6	10.37 [5.00–21.49] 11.78 [5.83–23.80]	0.804	2.78 [1.83–4.24] 3.78 [2.04–7.00]	0.419
Geographic origin	American: 4 Other: 5	13.37 [9.25–19.32] 9.75 [4.02–23.60]	0.518	3.34 [2.41–4.61] 3.39 [1.48–7.77]	0.971

these two dimensions [23, 52, 53]. We also believe that better understanding of these dimensions in studies related to predictive biomarkers or scoring systems for stratified medicine can improve the transfer of such tools into clinical practice.

From a methodological point of view, the large majority of papers in stratified medicine have focused on the evaluation of prognostic capacities, mainly by estimating area under ROC curves. In contrast, our results illustrate the necessity for a paradigm shift towards the evaluation of clinical usefulness, by estimating time-dependent utility function for instance [24].

Alternative decision-making approaches have been proposed that incorporate the consequences of clinical choices when assessing the usefulness of diagnostic tests or of prediction models. For instance, the widely used Decision Curve Analysis (DCA) computes the net benefit of a medical decision as the difference between its benefits and harms [54]. In DCA, the benefits and harms are obtained by asking the physicians for the probability threshold, i.e. the rate at which the patients are willing to trade-off between the harms and the benefits of a medical decision. A decision curve then plots the net benefit of a decision against the net benefits of "treating all" or "treating none" for the various possible values of the probability threshold. Although DCA may be easier to implement compared to our QALY-based approach, the two approaches may not lead to similar conclusions as they differ importantly. DCA relies on the physicians' prediction about the value of outcome [55], whereas patients' utility scores come from those experiencing (of having experienced) the health conditions under consideration. There is no guarantee that the physicians' judgments accurately reflect the views of the patients. Several studies found differences between the preferences of the patients and physicians for various conditions [56], especially in prostate cancer [57]. Although our QALY-based approach may be more in line with patients' preferences than DCA, the assessment of utility scores raises practical difficulties and published utility values are quite heterogeneous. Nevertheless, we believe that addressing these difficulties in future research may offer a more promising way to develop patient-centered approaches of medical decision making rather than relying solely on physicians' judgements.

It appears important to underline that the concept of clinical utility is multidimensional. It cannot be summarized into a single measure, whether it is QALYs or DCA. As explained by Smart [13], one can judge the concept of utility by using four components: appropriate, accessible, practicable, and acceptable. QALYs or DCA mainly concern the first one.

Our study nevertheless has some limitations. First, we identified a significant heterogeneity between the studies retained in our meta-analysis. We explored the possible causes of this heterogeneity, but we were not able to identify any significant reason. Second, the estimations of the SROCt curve [21] and the time-dependent utility function [24] both depend on many parametric assumptions. As with any meta-analysis on aggregated data, the corresponding limitation being listed by Lyman and Kuderer [58], it would have been preferable to collect the individual data necessary for such an analysis. To limit this issue, we have also non-parametrically adapted the two estimators after having reconstructed individual data (data not shown). The corresponding results were similar to the ones described in the main text. Third, numerous hypotheses are required to define the potential consequences of each alternative treatment (AM, RP and RHT) in terms of both quantity and quality of life. We attempted to be as realistic as possible in defining these hypotheses on data recently published in this field. Fourth, we did not study RP (for medium-risk patients) and RP + RHT (for high-risk patients). The reason is that it would require information about the removed tumor that is not included in the CAPRA score. Cooperberg et al. proposed the CAPRA-S score [59], as an extension of the CAPRA score that consider information about the tumor (surgical margin, seminal vesicle invasion, extracapsular extension, lymph node invasion). However, an analysis based on the CAPRA-S would be beyond the scope of our paper. Fifth, we separately investigated the discriminative capacities and the calibration of the CAPRA score. It would have been relevant to also estimate the global prognostic performance of the score, for instance by estimating the Index of Prognostic Accuracy,

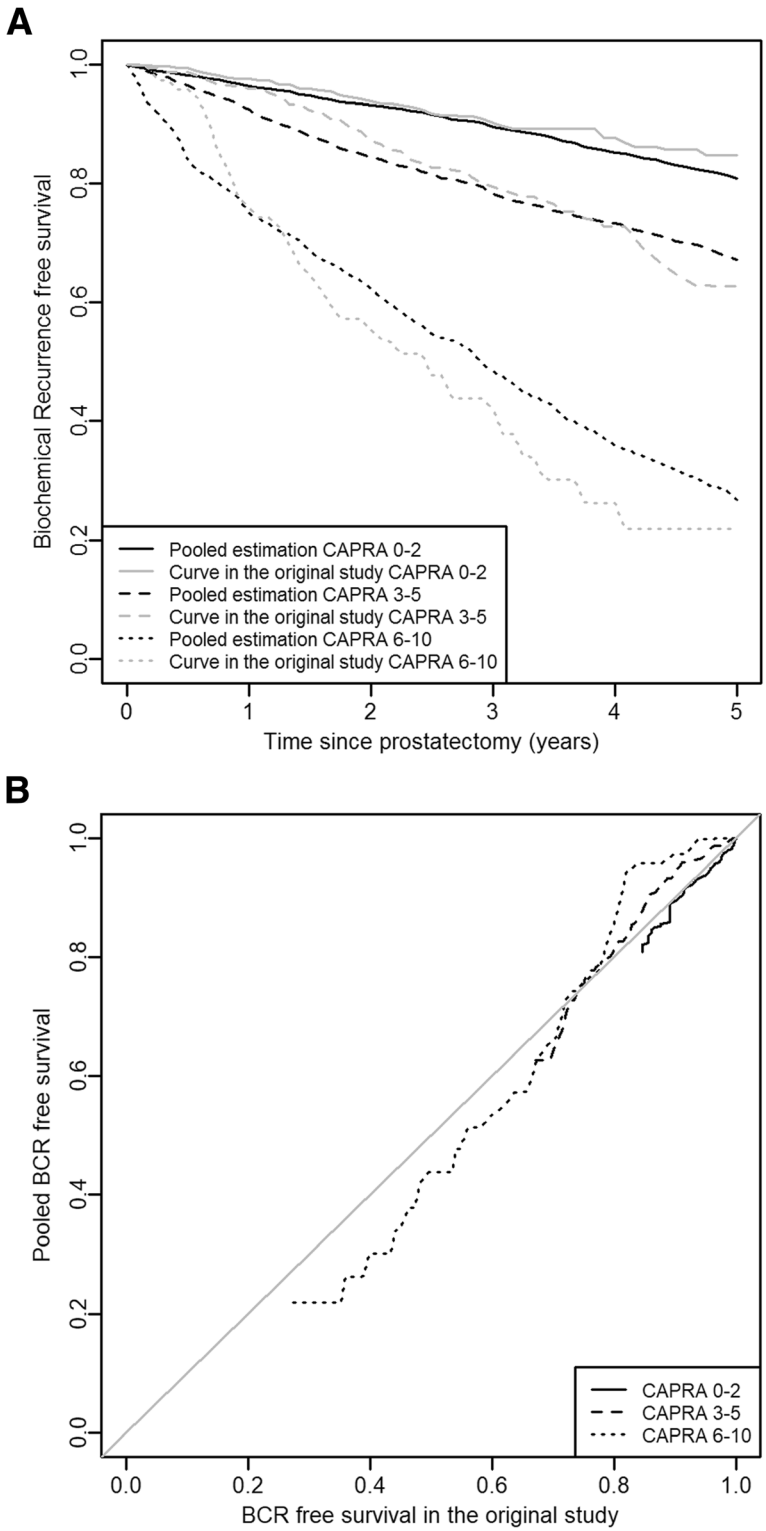


Fig. 3 Calibration plots. **a** The pooled survival curves according to CAPRA-based strata compared to the survival curves from the original study. **b** Pooled survival versus observed survival in the original study

as recently proposed by Kattan and Gerds [60]. Nevertheless, the corresponding pooled estimation from a published study does not exist to our knowledge and constitutes interesting perspective for future methodological developments.

Conclusions

Our study validated the prognostic capacities of the CAPRA score. Nevertheless, based on a QALYs maximi zation approach, our results challenged its clinical usefulness to beneficially stratify patients between three common therapeutic strategies: RHT for high-risk, RP for intermediate-risk, and AM for low-risk patients. Our results highlight the confusion between prognostic capacities and clinical usefulness. It calls for a better consideration of this distinction and for a paradigm shift from clinical validity to clinical utility in studies related to stratified medicine.

Additional file

Additional file 1: Calculation of the utility scores for the different patient health and care states. The corresponding text explains the calculation of the utility scores for the different patient health and care states. (PDF 109 kb)

Abbreviations

AM: Active monitoring; AUC: Area under the SROCt Curve; BCR: BioChemical Recurrence; CAPRA: CAncer of the Prostate Risk Assessment; HR: Hazard ratio; PSA: Prostate-specific antigen; QALY: Quality-Adjusted Life-Year; RHT: Radio-Hormonal Therapy; RP: Radical Prostatectomy; SROCt: Time-dependent Summary ROC

Acknowledgements

Not applicable.

Funding

This work was supported by the Cancer National Institute (INCa, Project MAP-MARKER, Decision N°2013–137). The INCa had no role in designing or editing this study.

Availability of data and materials

The datasets extracted from survival published survival curves and used during the current study are available from the corresponding author on reasonable request or online at the following address: https://www.labcom-risca.com/capra

Authors' contributions

ML participated in the statistical analyses and writing the manuscript. HM participated in data extraction and statistical analyses. PT participated in the utility collection and writing the paper. SS provided clinical expertise and participated in writing the paper. ED participated writing the paper and analyzing the data. YF supervised this work, participated writing the paper and analyzing the data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]SPHERE (methodS in Patient-centered outcomes & HEalth ResEarch) U1246, INSERM, Nantes University, Tours University, Nantes, France. [2]Division of Clinical Epidemiology and Aging Research, Heidelberg University, Heidelberg, Germany. [3]Department of Radiotherapy, Institut de Cancérologie de l'Ouest René Gauducheau, Saint Herblain, France. [4]INSERM UMR892, Nantes University, Nantes, France. [5]Nantes University Hospital, Nantes, France. [6]IRS2, SPHERE U1246, 22 boulevard Bénoni Goullin, 44200 Nantes, France.

Received: 10 May 2018 Accepted: 21 December 2018
Published online: 07 January 2019

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin. 2009;59:225–49.
2. Paller CJ, Antonarakis ES. Management of biochemically recurrent prostate cancer after local therapy: evolving standards of care and new directions. Clin Adv Hematol Oncol HO. 2013;11:14–23.
3. Antonarakis ES, Feng Z, Trock BJ, Humphreys EB, Carducci MA, Partin AW, et al. The natural history of metastatic progression in men with prostate-specific antigen recurrence after radical prostatectomy: long-term follow-up. BJU Int. 2012;109:32–9.
4. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA. 1998;280:969–74.
5. Roach M, Lu J, Pilepich MV, Asbell SO, Mohiuddin M, Terry R, et al. Four prognostic groups predict long-term survival from prostate cancer following radiotherapy alone on radiation therapy oncology group clinical trials. Int J Radiat Oncol Biol Phys. 2000;47:609–15.
6. Roach M, Weinberg V, Nash M, Sandler HM, McLaughlin PW, Kattan MW. Defining high risk prostate cancer with risk groups and nomograms: implications for designing clinical trials. J Urol. 2006;176(6 Pt 2):S16–20.
7. Huang J, Vicini FA, Williams SG, Ye H, McGrath S, Ghilezan M, et al. Percentage of positive biopsy cores: a better risk stratification model for prostate cancer? Int J Radiat Oncol Biol Phys. 2012;83:1141–8.
8. Cooperberg MR, Pasta DJ, Elkin EP, Litwin MS, Latini DM, DuChane J, et al. The UCSF Cancer of the prostate risk assessment (CAPRA) score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. J Urol. 2005;173:1938–42.
9. Kim Y-J, Cho KH, Pyo HR, Lee KH, Moon SH, Kim TH, et al. Radical prostatectomy versus external beam radiotherapy for localized prostate cancer: comparison of treatment outcomes. Strahlenther Onkol Organ Dtsch Rontgengesellschaft Al. 2014.
10. Meurs P, Galvin R, Fanning DM, Fahey T. Prognostic value of the CAPRA clinical prediction rule: a systematic review and meta-analysis. BJU Int. 2013; 111:427–36.
11. UCSF-CAPRA Score for Prostate Cancer Risk. MDCalc. https://www.mdcalc.com/ucsf-capra-score-prostate-cancer-risk. Accessed 18 Sep 2018.
12. UCSF Department of Urology | Prostate Cancer Risk Assessment and the UCSF-CAPRA Score. https://urology.ucsf.edu/research/cancer/prostate-cancer-risk-assessment-and-the-ucsf-capra-score. Accessed 18 Sep 2018.
13. Smart A. A multi-dimensional model of clinical utility. Int J Qual Health Care. 2006;18:377–82.
14. Development Core Team R. R: a language and environment for statistical computing. Vienna, Austria; 2010. http://www.R-project.org/
15. Poisot T. The digitize package: extracting numerical data from scatterplots. R J. 2011;3:25–6.
16. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med. 1998;17:2815–34.
17. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–88.
18. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34.
19. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539–58.

20. Combescure C, Foucher Y, Jackson D. Meta-analysis of single-arm survival studies: a distribution-free approach for estimating summary survival curves with random effects. Stat Med. 2014;33:2521–37.
21. Combescure C, Daures JP, Foucher Y. A literature-based approach to evaluate the predictive capacity of a marker using time-dependent summary receiver operating characteristics. Stat Methods Med Res. 2016;25:674–85.
22. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health. 2009; 12:S5–9.
23. Foucher Y, Lorent M, Tessier P, Supiot S, Sébille V, Dantan E. A mini-review of quality of life as an outcome in prostate cancer trials: patient-centered approaches are needed to propose appropriate treatments on behalf of patients. Health Qual Life Outcomes. 2018;16:40.
24. Dantan E, Foucher Y, Lorent M, Giral M, Tessier P. Optimal threshold estimator of a prognostic marker by maximizing a time-dependent expected utility function for a patient-centered stratified medicine. Stat Methods Med Res. 2018;27:1847–59.
25. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013;13:152.
26. Bill-Axelson A, Holmberg L, Filén F, Ruutu M, Garmo H, Busch C, et al. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial. J Natl Cancer Inst. 2008;100:1144–54.
27. Bolla M, Van Tienhoven G, Warde P, Dubois JB, Mirimanoff R-O, Storme G, et al. External irradiation with or without long-term androgen suppression for prostate cancer with high metastatic risk: 10-year results of an EORTC randomised study. Lancet Oncol. 2010;11:1066–73.
28. Denham JW, Joseph D, Lamb DS, Spry NA, Duchesne G, Matthews J, et al. Short-term androgen suppression and radiotherapy versus intermediate-term androgen suppression and radiotherapy, with or without zoledronic acid, in men with locally advanced prostate cancer (TROG 03.04 RADAR): an open-label, randomised, phase 3 factorial trial. Lancet Oncol. 2014;15:1076–89.
29. Horwitz EM, Bae K, Hanks GE, Porter A, Grignon DJ, Brereton HD, et al. Ten-year follow-up of radiation therapy oncology group protocol 92-02: a phase III trial of the duration of elective androgen deprivation in locally advanced prostate cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2008;26:2497–504.
30. Koerber F, Waidelich R, Stollenwerk B, Rogowski W. The cost-utility of open prostatectomy compared with active surveillance in early localised prostate cancer. BMC Health Serv Res. 2014;14:163.
31. Hanmer J, Vanness D, Gangnon R, Palta M, Fryback DG. Three methods tested to model SF-6D health utilities for health states involving comorbidity/co-occurring conditions. J Clin Epidemiol. 2010;63:331–41.
32. Ara R, Brazier J. Estimating health state utility values for comorbidities. PharmacoEconomics. 2017;35:89–94.
33. Loeb S, Curnyn C, Walter D, Fagerlin A, Siebert U, Mühlberger N, et al. Health state utilities among contemporary prostate cancer patients on active surveillance. Transl Androl Urol. 2018;7:197–202.
34. Krahn MD, Bremner KE, Alibhai SMH, Ni A, Tomlinson G, Laporte A, et al. A reference set of health utilities for long-term survivors of prostate cancer: population-based data from Ontario, Canada. Qual Life Res. 2013;22:2951–62.
35. Jayadevappa R, Schwartz JS, Chhatre S, Wein AJ, Bruce Malkowicz S. Association between utility and treatment among patients with prostate cancer. Qual Life Res. 2010;19:711–20.
36. Avila M, Becerra V, Guedea F, Suárez JF, Fernandez P, Macías V, et al. Estimating preferences for treatments in patients with localized prostate Cancer. Int J Radiat Oncol Biol Phys. 2014.
37. Stewart ST, Lenert L, Bhatnagar V, Kaplan RM. Utilities for prostate cancer health states in men aged 60 and older. Med Care. 2005;43:347–55.
38. Shimizu H, Horimoto Y, Arakawa A, Sonoue H, Kurata M, Kosaka T, et al. Application of a 70-gene expression profile to Japanese breast Cancer patients. Breast Care Basel Switz. 2015;10:118–22.
39. Cooperberg MR, Freedland SJ, Pasta DJ, Elkin EP, Presti JC, Amling CL, et al. Multiinstitutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. Cancer. 2006;107:2384–91.
40. Ishizaki F, Hoque MA, Nishiyama T, Kawasaki T, Kasahara T, Hara N, et al. External validation of the UCSF-CAPRA (University of California, san Francisco, Cancer of the prostate risk assessment) in Japanese patients receiving radical prostatectomy. Jpn J Clin Oncol. 2011;41:1259–64.
41. Loeb S, Carvalhal GF, Kan D, Desai A, Catalona WJ. External validation of the cancer of the prostate risk assessment (CAPRA) score in a single-surgeon radical prostatectomy series. Urol Oncol. 2012;30:584–9.
42. May M, Knoll N, Siegsmund M, Fahlenkamp D, Vogler H, Hoschke B, et al. Validity of the CAPRA score to predict biochemical recurrence-free survival after radical prostatectomy. Results from a european multicenter survey of 1,296 patients. J Urol. 2007;178:1957–62 discussion 1962.
43. Zhao KH, Hernandez DJ, Han M, Humphreys EB, Mangold LA, Partin AW. External validation of University of California, san Francisco, Cancer of the prostate risk assessment score. Urology. 2008;72:396–400.
44. Budaus L, Isbarn H, Tennstedt P, Salomon G, Schlomm T, Steuber T, et al. Risk assessment of metastatic recurrence in patients with prostate cancer by using the Cancer of the prostate risk assessment score: results from 2937 European patients. BJU Int. 2012;110:1714–20.
45. Seo WI, Kang PM, Chung JI. Predictive value of the cancer of the prostate risk assessment score for recurrence-free survival after radical prostatectomy in Korea: a single-surgeon series. Korean J Urol. 2014;55:321–6.
46. Yoshida T. Editorial comment to Japan Cancer of the prostate risk assessment for combined androgen blockade including bicalutamide: clinical application and validation. Int J Urol Off J Jpn Urol Assoc. 2013;20:714–5.
47. Tamblyn DJ, Chopra S, Yu C, Kattan MW, Pinnock C, Kopsaftis T. Comparative analysis of three risk assessment tools in Australian patients with prostate cancer. BJU Int. 2011;108(Suppl 2):51–6.
48. Lughezzani G, Lazzeri M, Larcher A, Lista G, Scattoni V, Cestari A, et al. Development and internal validation of a prostate health index based nomogram for predicting prostate cancer at extended biopsy. J Urol. 2012; 188:1144–50.
49. Hutchinson L. Closing the controversies gap in prostate cancer? Nat Rev Clin Oncol. 2014;11:299.
50. Chang AJ, Autio KA, Roach M III, Scher HI. High-risk prostate cancer[mdash] classification and therapy. Nat Rev Clin Oncol. 2014;11:308–23.
51. Klotz L, Emberton M. Management of low risk prostate cancer-active surveillance and focal therapy. Nat Rev Clin Oncol. 2014;11:324–34.
52. Donovan JL, Hamdy FC, Lane JA, Mason M, Metcalfe C, Walsh E, et al. Patient-reported outcomes after monitoring, surgery, or Radiotherapy for Prostate Cancer. N Engl J Med. 2016.
53. Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate Cancer. N Engl J Med. 2016;375:1415–24.
54. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak Int J Soc Med Decis Mak. 2006;26:565–74.
55. Traeger AC, Hübscher M, McAuley JH. Understanding the usefulness of prognostic models in clinical decision-making. J Physiother. 2017;63:121–5.
56. Montgomery AA, Fahey T. How do patients' treatment preferences compare with those of clinicians? Qual Health Care QHC. 2001;10(Suppl 1):i39–43.
57. Tentori K, Pighin S, Divan C, Crupi V. Mind the gap: physicians' assessment of patients' importance weights in localized prostate cancer. PLoS One. 2018;13:e0200780.
58. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC Med Res Methodol. 2005;5:14.
59. 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.
60. Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res. 2018;2:7.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

