Original Article

External validation of a risk model predicting failure of salvage focal ablation for prostate cancer

Alexander Light[1,2], Max Peters[20], Deepika Reddy[1,2], Abi Kanthabalan[3], Marjorie Otieno[3], Menelaos Pavlou[4], Rumana Omar[4], Sola Adeleke[6,7], Francesco Giganti[3,8], Chris Brew-Graves[5], Norman R. Williams[3], Amr Emara[14], Athar Haroon[12,13], Arash Latifoltojar[5,15], Harbir Sidhu[5,8], Alex Freeman[9], Clement Orczyk[3,10], Ashok Nikapota[16], Tim Dudderidge[17], Richard G. Hindley[14], Jaspal Virdi[18], Manit Arya[2], Heather Payne[9], Anita V. Mitra[11], Jamshed Bomanji[12], Mathias Winkler[1,2], Gail Horan[19], Caroline Moore[3,10], Mark Emberton[3,10], Shonit Punwani[5,8], Hashim U. Ahmed[1,2,3] and Taimur T. Shah[1,2,3]

[1]Imperial Prostate, Department of Surgery and Cancer, Imperial College London, [2]Imperial Urology, Charing Cross Hospital, Imperial College Healthcare NHS Trust, [3]Division of Surgery and Interventional Sciences, [4]Department of Statistical Science, [5]Division of Medicine, Faculty of Medicine, University College London, [6]Department of Oncology, Guy's and St Thomas' NHS Foundation Trust, [7]School of Cancer and Pharmaceutical Sciences, King's College London, [8]Department of Radiology, [9]Department of Histopathology, [10]Department of Urology, [11]Department of Oncology, University College London Hospital NHS Foundation Trust, [12]Institute of Nuclear Medicine, University College London Hospitals NHS Foundation Trust, [13]Department of Nuclear Medicine, St Bartholomew's Hospital, Barts Health NHS Trust, London, [14]Department of Urology, Basingstoke and North Hampshire Hospital, Hampshire Hospitals NHS Foundation Trust, Basingstoke, [15]Department of Radiology, Royal Marsden NHS Foundation Trust, Surrey, [16]Sussex Cancer Centre, Royal Sussex County Hospital, University Hospitals Sussex NHS Foundation Trust, Brighton, [17]Department of Urology, University Hospital Southampton NHS Foundation Trust, Southampton, [18]Department of Urology, The Princess Alexandra Hospital NHS Trust, Harlow, [19]Department of Oncology, Queen Elizabeth Hospital, The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, King's Lynn, UK, and [20]Department of Radiation Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

H.U.A. and T.T.S. joint senior authors.

Objectives

To externally validate a published model predicting failure within 2 years after salvage focal ablation in men with localised radiorecurrent prostate cancer using a prospective, UK multicentre dataset.

Patients and methods

Patients with biopsy-confirmed ≤T3bN0M0 cancer after previous external beam radiotherapy or brachytherapy were included from the FOcal RECurrent Assessment and Salvage Treatment (FORECAST) trial (NCT01883128; 2014–2018; six centres), and from the high-intensity focussed ultrasound (HIFU) Evaluation and Assessment of Treatment (HEAT) and International Cryotherapy Evaluation (ICE) UK-based registries (2006–2022; nine centres). Eligible patients underwent either salvage focal HIFU or cryotherapy, with the choice based predominantly on anatomical factors. Per the original multivariable Cox regression model, the predicted outcome was a composite failure outcome. Model performance was assessed at 2 years post-salvage with discrimination (concordance index [C-index]), calibration (calibration curve and slope), and decision curve analysis. For the latter, two clinically-reasonable risk threshold ranges of 0.14–0.52 and 0.26–0.36 were considered, corresponding to previously published pooled 2-year recurrence-free survival rates for salvage local treatments.

Results

A total of 168 patients were included, of whom 84/168 (50%) experienced the primary outcome in all follow-ups, and 72/168 (43%) within 2 years. The C-index was 0.65 (95% confidence interval 0.58–0.71). On graphical inspection, there was close agreement between predicted and observed failure. The calibration slope was 1.01. In decision curve analysis, there was incremental net benefit vs a 'treat all' strategy at risk thresholds of ≥0.23. The net benefit was therefore higher across the majority of the 0.14–0.52 risk threshold range, and all of the 0.26–0.36 range.

Conclusion

In external validation using prospective, multicentre data, this model demonstrated modest discrimination but good calibration and clinical utility for predicting failure of salvage focal ablation within 2 years. This model could be reasonably

used to improve selection of appropriate treatment candidates for salvage focal ablation, and its use should be considered when discussing salvage options with patients. Further validation in larger, international cohorts with longer follow-up is recommended.

Keywords

ablation, cryotherapy, failure, focal therapy, high-intensity focussed ultrasound, prediction model, prostate cancer, radiotherapy, recurrence, salvage

Introduction

Radiotherapy (RT) is a common and effective prostate cancer treatment for many patients, with >12 000 UK men undergoing external-beam RT (EBRT) each year [1]. However, ~10% with intermediate- or high-risk disease will develop recurrence localised to the prostate over long-term follow-up, an event independently predictive of metastasis and cancer-specific death [2]. Patients with localised radiorecurrence are typically offered surveillance or non-curative androgen-deprivation therapy (ADT). Whole-gland salvage treatments are offered by some centres to highly-selected patients but confer high rates of toxicity. Salvage radical prostatectomy, e.g., leads to erectile dysfunction in nearly all, urinary incontinence in 80%, and rectal injury in 5%–10% [3,4]. An emerging alternative is salvage focal ablation. Encompassing treatments like high-intensity focussed ultrasound (HIFU) and cryotherapy, this targets the recurrent lesion(s) alone. Preliminary data suggest this provides good early disease control with reduced toxicity [5,6].

Despite this, optimal patient selection for focal ablation remains unknown. Certainly, the risk of treatment failure should be central to these decisions. In 2018, Peters et al. [7] developed and internally validated a model for predicting failure after salvage focal HIFU. To our knowledge, this is the only published model predicting failure after salvage focal ablation. This study aimed to externally validate this model using prospective, UK multicentre data from a cohort study and two national registries.

Patients and Methods

Validation Cohort

Patients were enrolled either within the FOcal RECurrent Assessment and Salvage Treatment (FORECAST) trial (NCT01883128), or the HIFU Evaluation and Assessment of Treatment (HEAT) and International Cryotherapy Evaluation (ICE) UK national registries [6,8,9]. All underwent salvage focal HIFU or cryotherapy after previous EBRT and/or low/high dose rate brachytherapy with or without (neo)adjuvant ADT. For this analysis, only patients with ≤T3bN0M0 radiorecurrent disease were included, matching the inclusion criteria of the original model [7].

Forecast

Between 2014 and 2018, 181 patients were prospectively enrolled to six UK centres who had biochemical failure defined by rising PSA levels post-RT [6]. Those taking ADT within 6 months of enrolment, with a PSA doubling time of ≤3 months, with a total PSA level of ≥20 ng/mL, unable to have an MRI, or with previous salvage treatment were ineligible.

Following ^{18}F-choline positron emission tomography (PET)/CT and ^{99m}Tc methylene diphosphonate bone scan, patients underwent prostate multiparametric MRI (mpMRI) followed by transperineal mpMRI-targeted and template mapping biopsies [10]. Eligible patients were offered either salvage HIFU or cryotherapy, according to disease location, alongside other options like salvage prostatectomy or observation, as defined by a multidisciplinary team meeting. Cryotherapy was used for anterior tumours, larger tumours with an anterior–posterior distance of >3.5 cm, and prostates with calcifications or previous brachytherapy seeds. All other patients with peripheral zone or posterior tumours underwent HIFU.

In the 93 patients who underwent focal ablation, PSA measurements were taken postoperatively at 1, 3, 6, 9, and 12 months, then every 6 months. A prostate mpMRI was also routinely performed at 12 months. Any further imaging or biopsy were ordered based on clinical judgement and were not protocol mandated.

The HEAT and ICE Registries

Between 2006 and 2022, 292 patients with radiorecurrence undergoing salvage HIFU or cryotherapy from nine UK centres were prospectively enrolled into HEAT and ICE [8,9]. Radiorecurrence was based on a rising PSA meeting Phoenix criteria, which triggered re-staging investigations, comprising bone scan, CT, ^{18}F-choline PET/CT, or prostate-specific membrane antigen (PSMA) PET/CT dependent on local practice. Patients then underwent mpMRI with systematic and mpMRI-targeted biopsy.

Salvage focal ablation was offered to patients with non-metastatic disease requiring a maximum of 75% ablation of the gland. The decision of which energy to use was made locally; however, generally HIFU was used for posterior disease and cryotherapy for anterior or T3b disease.

Postoperatively, subsequent PSA measurements, imaging, and biopsy were ordered based on local practice and protocols.

Outcome

The primary outcome was treatment failure as defined by the original model [7]. This was a composite of any of the following: biochemical failure (PSA value ≥2 ng/mL above nadir), localised/distant disease on imaging (prostate mpMRI, PET/CT, bone scan), positive repeat biopsy, initiation of systemic treatment (ADT, chemotherapy), or cancer-specific death. Patients who did not fail were censored at the date of their latest appointment or investigation. Neither clinicians nor the study team were blinded to collection of predictor or outcome data.

Statistical Analysis

The model by Peters et al. [7] comprises a score developed from a multivariable Cox regression model. Model variables measured at the time of radiorecurrence diagnosis, and their individual coefficients, were: Gleason score (Gleason 7: −0.083; Gleason 8–10: 0.48), radiological T-stage (T3: 0.314), PSA (ng/mL; 0.042), prostate volume (mL; 0.007), and disease-free survival (DFS) interval (months; −0.007). The DFS interval was the duration between finishing primary treatment and the mpMRI assessing for radiorecurrence. If mpMRI date was unavailable, the biopsy date was used. As this variable was measured in months, if relevant dates were only available in year format rather than year and month, then this datapoint was considered missing so as to avoid inappropriately biasing this variable. For all variables, missing data were considered missing at random, and derived using multiple imputation by chain equations with 20 iterations and 1000 re-samples using the *mice* R package. Model variables were used for imputation in addition to the binary failure outcome and the Nelson–Aalen estimate of the cumulative hazard function [11]. Missing data were imputed for PSA for seven patients (4%), Gleason score for 18 (11%), prostate volume for 46 (27%), and DFS interval for 61 (36%).

To calculate the risk score, variable coefficients were multiplied by 10 then multiplied by their respective value, and then summated. An additional 10 points were then added to obtain positive sum scores. For example, for a patient with radiorecurrent Gleason 8, T3a cancer with a PSA level of 5 ng/mL, a prostate volume of 40 mL and 72 month DFS interval, the risk score would be calculated as follows:

$$
\begin{aligned}
&(0.48 \times 10)\,[\text{Gleason 8}] + (0.314 \times 10)\,[\text{T3 stage}] \\
&\quad + (5 \times 0.042 \times 10)\,[\text{PSA 5 ng/mL}] \\
&\quad + (40 \times 0.007 \times 10)\,[\text{volume 40 mL}] \\
&\quad + (72 \times -0.007 \times 10)[\text{disease} \\
&\quad - \text{free survival interval 72 months}] + 10 \\
&= 4.8 + 3.14 + 2.1 + 2.8 - 5.04 + 10 \\
&= 17.8 \text{ points}
\end{aligned}
$$

A univariable Cox regression model was fitted using calculated risk score in order to predict failure. Model performance was assessed at 2 years post-ablation, with performance measures estimated in each imputed dataset and then summarised using Rubin's rules [12]. Discrimination was evaluated using the concordance index (C-index). Calibration was assessed graphically through plotting predicted vs observed failure, and through calculation of the calibration slope.

Decision curve analysis was also performed to determine clinical utility [13]. Here, net benefit (*y*-axis) is plotted against risk threshold (*x*-axis). Risk threshold refers to clinician preferences in regard to offering salvage focal ablation, taking account of its benefits vs harms. Lower risk thresholds reflect clinicians who are more concerned with missing the benefits of salvage focal ablation, that is the opportunity to treat any recurrence successfully. This therefore represents a low threshold for offering treatment. Higher risk thresholds reflect clinicians who are more concerned regarding the harms of salvage focal ablation, that is fewer, better selected patients should be treated in order to minimise any harms; thus these clinicians have a high threshold for offering treatment. A given risk threshold is defined as the minimum probability of failure at which salvage focal ablation would be warranted. Model net benefit, which takes into account both discrimination and calibration, is a combination of model-predicted false positives subtracted from true positives weighted against a given risk threshold.

In decision curve analysis, model-based decision making on whether to offer salvage focal ablation is compared against strategies of treating all patients and treating no patients. The model with the highest net benefit across a clinically-reasonable range of risk thresholds has the greatest clinical utility and can be recommended for use. As the reference strategy in this scenario is treating all patients, net benefit can also be expressed in terms of true negatives, equivalent to the number of salvage focal ablation procedures that can be avoided. These patients who avoid a procedure reflect a high predicted risk of failure that may instead warrant whole-gland or multi-modal treatment strategies.

To determine a clinically-reasonable range of risk thresholds, we considered a 2021 systematic review that calculated pooled 2-year recurrence-free survival (RFS) rates for six salvage local treatments post-RT. The lowest rate was reported for focal and whole-gland HIFU (54%, 95% CI 48%–60%), and the highest rate for low dose-rate brachytherapy (81%, 95% CI 74%–86%) [4]. Taking the lowest and highest bounds of these two 95% CIs, a 48%–86% 2-year RFS rate is equivalent to a 14%–52% 2-year recurrence rate. This risk threshold range of 0.14–0.52 was the first clinically-reasonable range of risk thresholds considered. We also considered a second range as determined by the 2-year RFS rate of salvage radical prostatectomy (69%, 95% CI 64%–74%), a risk threshold range of 0.26–0.36.

Fig. 1 Flow chart detailing the exclusion and inclusion process of men in this study.
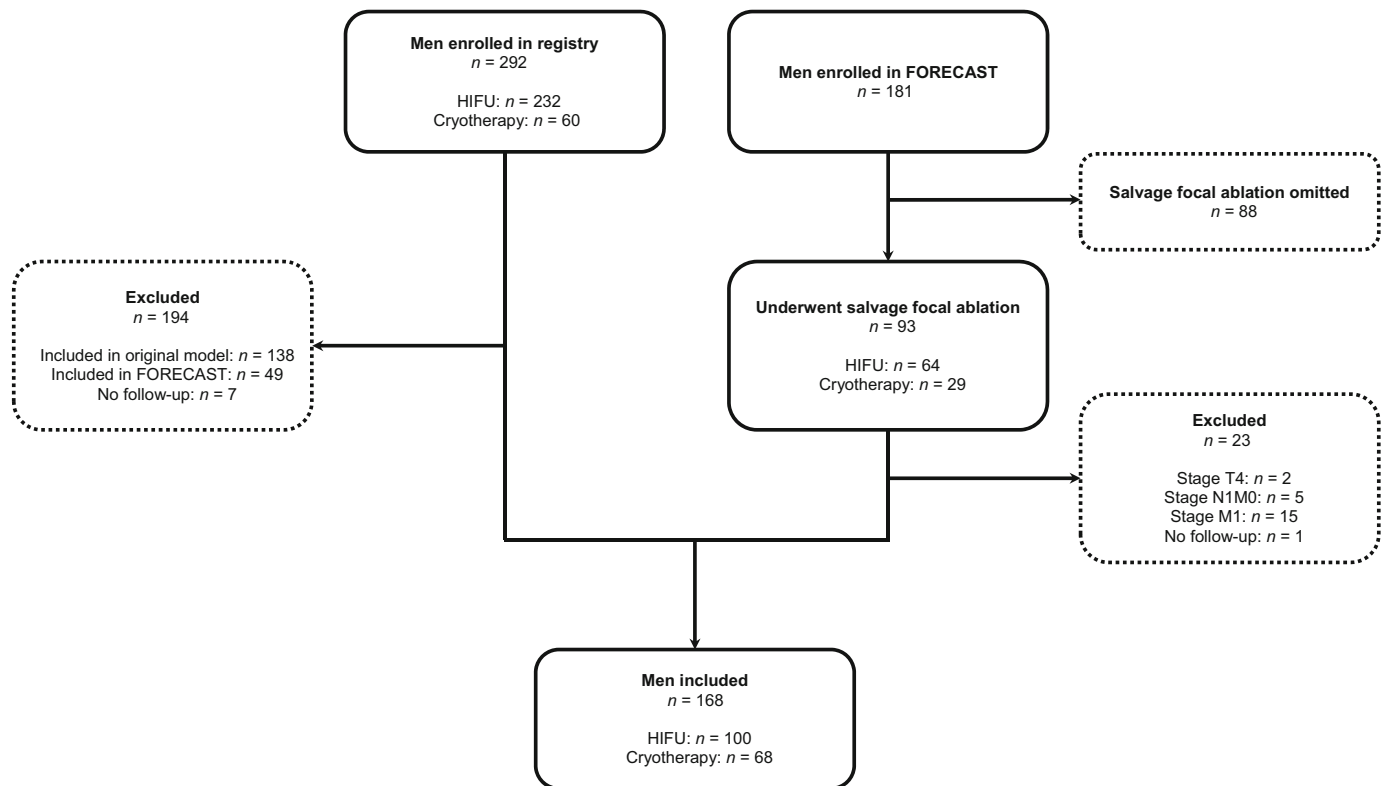
Next, patients were categorised into three risk groups described by the original study, which were created based on 4-year failure-free survival proportions [7]. These were: Group 1, score ≤7 (best prognosis); Group 2, scores >7 and ≤15; and Group 3, scores >15 (worst prognosis). Failure-free survival distributions were plotted using Kaplan–Meier curves and compared using log-rank tests adjusted for multiple comparisons using the Benjamini–Hochberg method [14].

Validation was performed using all patients who underwent salvage focal ablation. As subgroup analyses, discrimination and calibration were then estimated separately by ablation energy and data source.

Analyses were performed using R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $P < 0.05$.

Sample Size

The minimum sample size to provide precise estimation of the calibration slope was calculated as per Riley et al. [15] (Appendix S1). As reported by the original study, for a C-index of 0.64 from internal validation and an estimated survival probability of 0.54 at 2 years, our cohort sample size of 164 would provide a 95% CI of 0.77–1.23 for a calibration slope of 1 [7]. This assumes no censoring prior to the 2 year

timepoint. Censoring prior to 2 years was not reported by the original study; however, in our cohort this was 35%. Therefore, assuming a 35% censor rate prior to 2 years, a sample size of 164 would give 95% CI of 0.05–1.95. For a target 95% CI of 0.9–1.1, as recommended by Riley et al. [15], a minimum sample size of 8000 and 12 500 would be required assuming no censoring and assuming a 35% censor rate by 2 years, respectively.

Results

From 493 patients considered, 168 were eligible (Fig. 1). In all, 100/168 (60%) underwent salvage focal HIFU, and 68/168 (40%) underwent salvage focal cryotherapy. Table 1 shows the patients' clinicopathological data split by ablation energy, with Table S1 showing combined data. Table S2 shows these data vs the original development cohort of 150 patients [7]. The characteristics appeared generally comparable, although this external validation cohort comprised a greater proportion of brachytherapy patients (difference 17% [20% vs 3%], Fisher's exact test $P < 0.001$). Furthermore, the original development cohort comprised purely HIFU-treated patients.

In this external validation cohort, 84/168 patients (50%) experienced the primary failure outcome in all follow-ups (HIFU, $n = 50$; cryotherapy, $n = 34$) and 72/168 patients

Table 1 Patients' characteristics, both at diagnosis and at time of enrolment in FORECAST or in the HEAT and ICE registries, split by ablation energy.

Characteristic	HIFU, *n* = 100	Cryotherapy, *n* = 68	*P*[*]
At diagnosis			
Age, years, median (IQR)	62.0 (58.0, 66.0)	64.0 (60.8, 69.0)	0.01
Unknown, *n*	20	12	
PSA level, ng/mL, median (IQR)	13.0 (7.7, 23.2)	10.9 (7.8, 22.2)	0.8
Unknown, *n*	20	20	
Gleason score, *n* (%)			
≤6	29 (36)	21 (41)	0.8
7	40 (49)	22 (43)	
≥8	12 (15)	8 (16)	
Unknown	19	17	
At trial or registry enrolment			
EBRT use, *n* (%)	88 (88)	52 (78)	0.07
Unknown	0	1	
Brachytherapy use, *n* (%)	13 (13)	21 (31)	0.004
Unknown	0	1	
Neoadjuvant/adjuvant hormone use, *n* (%)	69 (80)	15 (83)	0.8
Unknown	14	50	
Age, years, median (IQR)	69.0 (66.2, 73.0)	72.0 (68.0, 76.0)	0.02
Unknown, *n*	6	6	
DFS interval, months, median (IQR)	86 (61, 116)	87 (53, 118)	0.8
Unknown, *n*	40	21	
PSA level, ng/mL, median (IQR)	4.7 (2.5, 8.0)	5.1 (3.0, 7.2)	0.9
Unknown, *n*	3	4	
Prostate volume, mL, median (IQR)	22.0 (19.0, 30.0)	24.0 (20.0, 34.4)	0.1
Unknown, *n*	19	27	
MRI stage, *n* (%)			
T1/2	74 (79)	45 (82)	0.8
T3	1 (1.1)	0 (0)	
T3a	11 (12)	4 (7.3)	
T3b	8 (8.5)	6 (11)	
Unknown	6	13	
Total biopsy cores, median (IQR)	29 (22, 40)	23 (15, 36)	0.1
Unknown, *n*	3	7	
Positive biopsy cores, median (IQR)	5 (3, 9)	6 (3, 8)	0.6
Unknown, *n*	4	7	
Grade group, *n* (%)			
1	3 (3.3)	3 (4.9)	0.03
2	26 (29)	15 (25)	
3	37 (41)	19 (31)	
4	16 (18)	6 (9.8)	
5	9 (9.9)	18 (30)	
Unknown or irradiation effect	9	7	
MCCL, mm, median (IQR)	7 (4, 10)	8 (5, 10)	0.2
Unknown, *n*	11	8	

[*]*Wilcoxon rank-sum test; Pearson's chi-squared test; Fisher's exact test. MCCL, maximum cancer core length.*
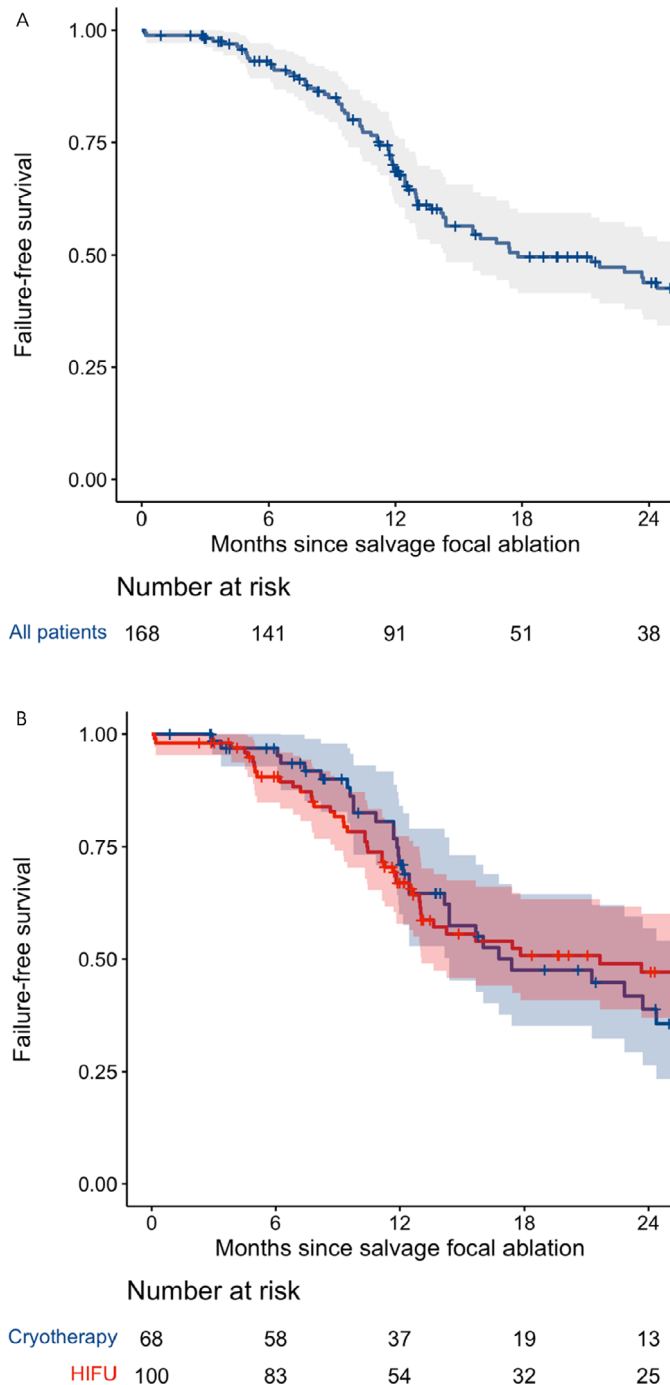
(43%) experienced this within 2 years (HIFU, *n* = 43; cryotherapy, *n* = 29). Specifically, 34/72 (47%) experienced biochemical failure, 31/72 (43%) had positive re-imaging, four of 72 (6%) had a positive re-biopsy, and three of 72 (4%) started systemic treatment. In all cases, ADT was started prior to any other systemic treatment. No patient experienced cancer-specific mortality within 2 years, although in all follow-ups this affected two patients at times of 8.65 and 10.62 years post-salvage focal ablation. Fig. 2 displays Kaplan–Meier curves illustrating failure-free survival distribution for all included patients and stratified by energy. The median (interquartile range [IQR]) time-to-failure was 0.99 (0.71–1.35) years, and the median (IQR) follow-up for censored patients was 1.09 (0.68–2.21) years. Table S3 shows the number of patients reaching each individual outcome of the composite failure outcome in all follow-ups, with Fig. S1 showing corresponding Kaplan–Meier curves.

Model discrimination was modest (C-index 0.65, 95% CI 0.58–0.71). Calibration was good; however, with close agreement between predicted and observed failure on inspection of the calibration curve (Fig. 3). Furthermore, the calibration slope was 1.01.

In decision curve analysis, there was incremental net benefit using model-based decision-making compared to a 'treat all' strategy at risk thresholds ≥0.23 (Fig. 3) [13]. For risk

Fig. 2 Kaplan–Meier curves plotting failure-free survival distributions for all patients undergoing salvage focal ablation (**A**) and stratified by focal ablation energy (**B**).

higher risk thresholds within this range, suggesting that proportionally greater model benefit lies with clinicians who prefer to be more selective of patients. In this range, using the model to select treatment candidates vs treating all patients would lead to a 0%–9.5% reduction in salvage focal ablation procedures. Considering the second risk threshold range of 0.26–0.36, model-based decision making was the optimal strategy for all risk thresholds. However, on inspection, net benefit was only very marginally greater throughout this range. In this range, there would be a 0.7%–2.6% reduction in salvage focal ablation procedures performed.

Fig. 4 displays Kaplan–Meier curves for each of the three risk-score groups as detailed in the original development study [7]. In adjusted pairwise log-rank tests, there was a significant difference in failure-free survival distributions between groups 1 and 2 ($P < 0.001$), groups 1 and 3 ($P < 0.001$), and groups 2 and 3 ($P = 0.032$), reflecting the good calibration of the model in this cohort.

The C-index across the four subgroup analyses ranged from 0.60–0.68 (HIFU only: 0.64, 95% CI 0.55–0.72; cryotherapy only: 0.68, 95% CI 0.57–0.78; FORECAST only: 0.60, 95% CI 0.51–0.79; HEAT/ICE registry only: 0.67, 95% CI 0.58–0.77). There remained generally good agreement between predicted and observed failure in the cryotherapy-only subgroup (calibration slope 1.10), the FORECAST-only subgroup (calibration slope 0.88), and HEAT/ICE registry-only subgroup (calibration slope 1.10; Fig. S2). However, for HIFU patients, the model slightly overestimated failure at lower predictions, and underestimated it at higher predictions (calibration slope 0.98).

Discussion

Summary of Results

In this external validation, the 43% 2-year and 50% all follow-up failure rate demonstrated emphasises the need for an effective risk model to predict which patients are likely to fail treatment, and thus for whom offering salvage focal ablation may not be warranted. The multivariable risk model here demonstrated comparably modest discrimination to internal validation (0.65 vs 0.64, respectively), but with good calibration [7]. Furthermore, compared to a 'treat all' strategy, there was incremental net benefit for clinicians across the majority of risk thresholds in the range 0.14–0.52, corresponding to previously published pooled 2-year RFS rates of local salvage treatments [4]. Importantly, there was also incremental net benefit for all risk thresholds in the range 0.26–0.36, which corresponds to the pooled 2-year RFS rate of salvage radical prostatectomy from the same analysis. Certainly, therefore, this model does offer some utility to clinicians who want to
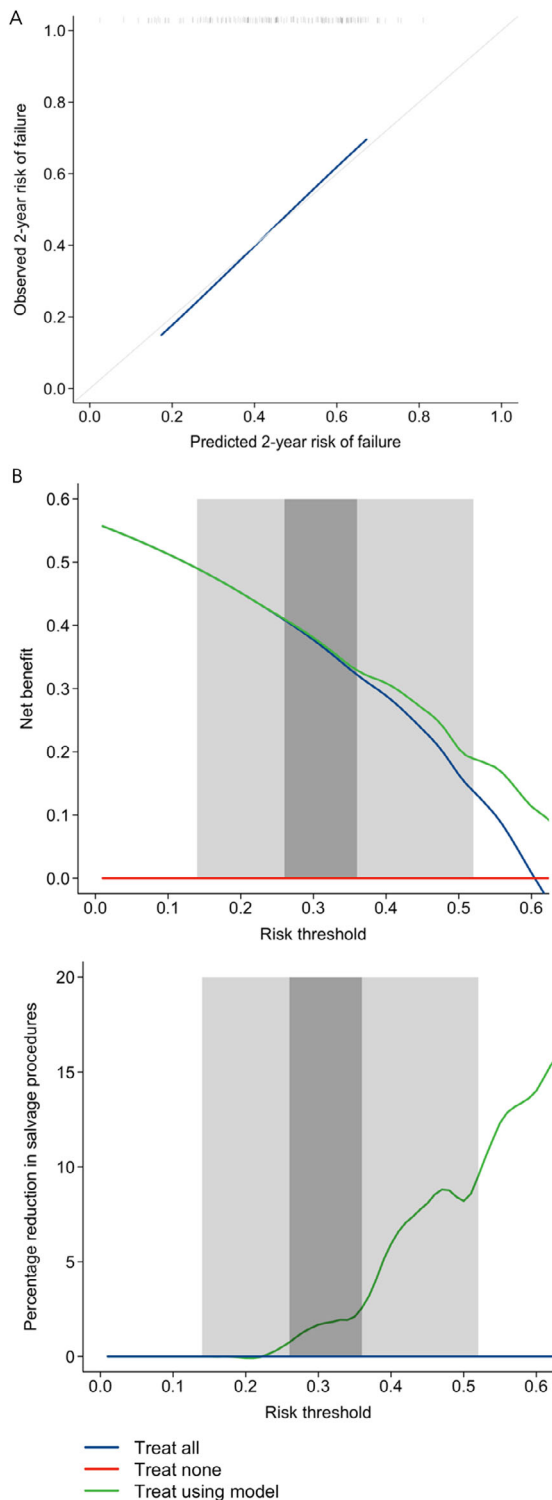
thresholds of <0.23, using the model to select patients for treatment had no benefit compared to offering treatment to all patients. Therefore, considering the risk threshold range of 0.14–0.52, model-based decision making represents the optimal strategy for the majority of these risk thresholds. In addition, a proportionally greater net benefit was observed at

Fig. 3 Calibration curve (**A**) and decision curve analysis (**B**) for model predictions of composite failure at 2 years post-salvage focal ablation for all included men undergoing salvage focal ablation. Calibration slope was 1.01. Decision curve analysis compares decision making to offer salvage focal ablation between model-based decision making and strategies of treating all patients and treating no patients. Plots demonstrating net benefit and percentage reduction in salvage focal ablation procedures are shown. Two clinically-reasonable ranges of risk threshold are highlighted: (i) 0.14–0.52 (light grey); and (ii) 0.26–0.36 (dark grey), based on previously-published pooled 2-year recurrence rates probabilities [4].

patient's risk score should be considered when discussing salvage options, and the risk groups detailed could be of use. Patients with higher risk predictions, e.g., Risk Group 3, can be more appropriately counselled and may benefit from alternative discussion of whole-gland treatments or multi-modal therapy. In contrast, patients with lower predictions, e.g., Risk Group 1, can be reassured regarding the early efficacy of salvage focal treatment. These risk groups could also be used to guide intensity of follow-up postoperatively.

It should be noted that HIFU and cryotherapy are deemed here to be complementary rather than competing treatments, with selection of either therapy based predominantly on anatomical factors. In subgroup analysis, the calibration in HIFU-treated patients was notably worse on inspection of the calibration curve compared to cryotherapy-treated patients, and this should be considered with clinical use of the model. This may be due to a higher proportion of the external validation HIFU-treated cohort having undergone previous brachytherapy compared to the development cohort (difference 11% [13% vs 3%], Fisher's exact test $P < 0.001$). The presence of brachytherapy seeds, typically, is a relative indication to treat with cryotherapy over HIFU [6]. Nonetheless, despite being developed in an exclusively HIFU-treated cohort, it is encouraging to see good model calibration for cryotherapy-treated patients in this validation.

The primary composite failure outcome is designed to reflect potential disease progression as reached via different scenarios [7]. Consequently, the next steps for patients experiencing this outcome are not uniform. Nonetheless, unless already performed, this may include any combination of local and/or whole-body re-imaging, re-biopsy, ongoing observation, further local salvage treatment, and/or commencement of ADT. Importantly, this model and its predicted outcome do not seek to recommend specific next steps; instead, for a patient that has met the composite failure outcome, this implies that further investigation and potentially treatment is indicated as decided by their clinician.
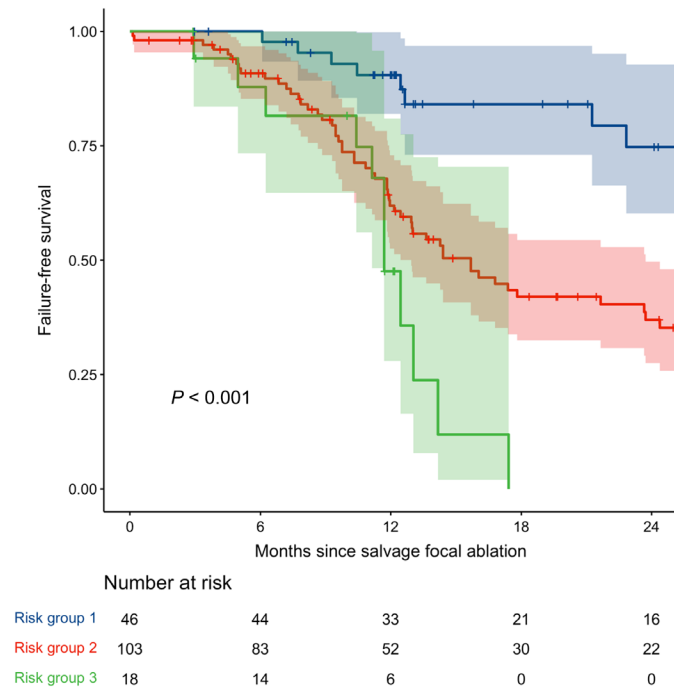
Context

Improving the management of radiorecurrence is an important but under-studied research need. Considering that

be more selective of treatment candidates for salvage HIFU and cryotherapy.

A nomogram to facilitate clinical use is detailed in Fig. 5, using the original model's coefficients. Calculation of a

Fig. 4 Kaplan–Meier curves curve plotting time-to-failure stratified by risk groups. This is plotted for all included men undergoing salvage focal ablation. Log-rank test demonstrated a significant difference in survival distributions between groups ($P < 0.001$).
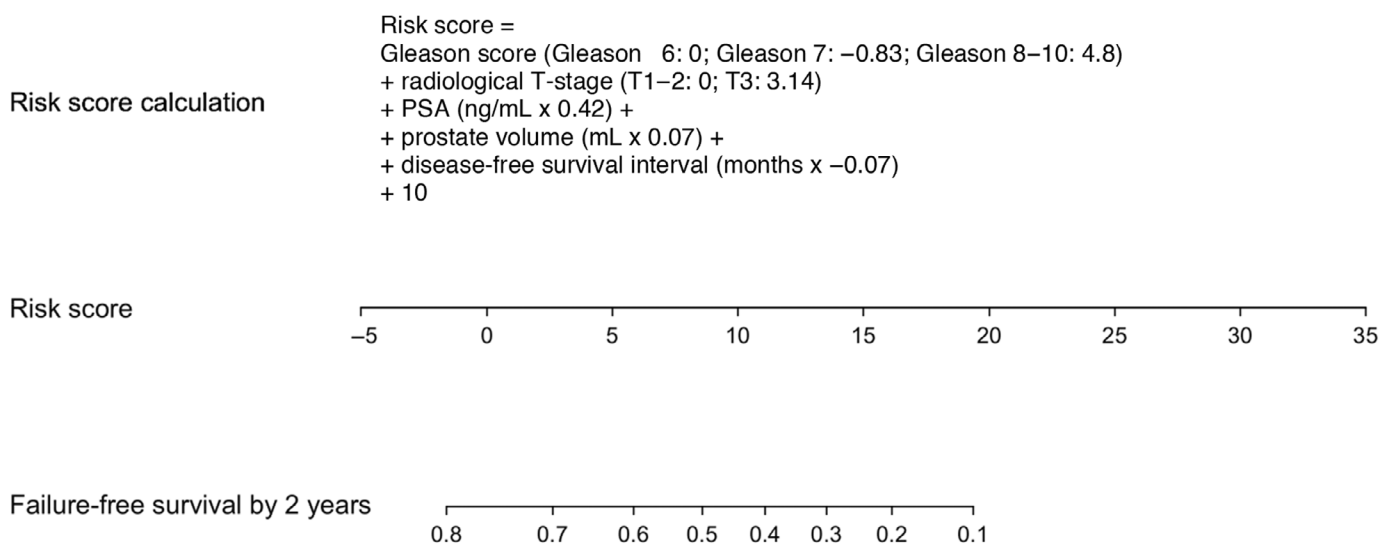

Number at risk

	0	6	12	18	24
Risk group 1	46	44	33	21	16
Risk group 2	103	83	52	30	22
Risk group 3	18	14	6	0	0

in the UK alone, >12 000 patients undergo EBRT for prostate cancer each year, ~20% will develop biochemical failure [1,16]. Furthermore, within 5 years of biochemical failure, 50% will develop distant metastases, and 20% will die from their cancer [16]. Overall, recurrence confined to the prostate affects 10% of patients with intermediate- and high-risk

disease and is independently predictive of metastasis and cancer-specific death [2]. It follows that preventing or delaying metastases and subsequent death in these patients through effective treatment of localised disease is therefore crucial.

At the point of biochemical failure, watchful waiting or non-curative ADT is typically offered. However, the latter carries bothersome side-effects like hot flushes and reduced libido, plus significant metabolic toxicity [17]. Furthermore, castrate-resistant disease develops after 2–3 years, requiring expensive second- and third-line systemic agents [18]. As an alternative, our group has previously shown that, based on transperineal template mapping biopsies, as many as three-quarters of patients with localised radiorecurrent disease may be anatomically suitable for salvage focal ablation [19]. In support of the salvage focal approach, FORECAST demonstrated that focal ablation provides good early cancer control with 66% progression-free survival and preserved urinary continence in 84% at 2 years follow-up [6]. This is also supported by a 2020 systematic review, concluding a 48%–72% 3-year DFS rate [5].

At present, few centres perform salvage therapy post-RT, and even fewer offer salvage focal therapy. Wider application of salvage focal treatments will require optimisation of two key areas: (i) accurate diagnosis and localisation of radiorecurrent disease; and (ii) patient selection for salvage treatment. FORECAST addressed the former, showing that radiorecurrent cancer is prevalent in those with a rising PSA (80%), and that mpMRI followed by both systematic and targeted biopsies is important for detecting this [6]. The model by Peters et al. [7] and the present external validation, addresses the second area, showing that short-term

Fig. 5 Method of risk score calculation, presented with a nomogram presenting probability of failure-free survival by 2 years corresponding to the range of possible risk scores.

Risk score calculation

Risk score =
Gleason score (Gleason 6: 0; Gleason 7: −0.83; Gleason 8–10: 4.8)
+ radiological T-stage (T1−2: 0; T3: 3.14)
+ PSA (ng/mL x 0.42) +
+ prostate volume (mL x 0.07) +
+ disease-free survival interval (months x −0.07)
+ 10

Risk score

Failure-free survival by 2 years

oncological outcome post-ablation can be predicted with reasonable performance, and that higher- vs lower-risk treatment candidates can be distinguished.

To our knowledge, this model is the only published tool predicting failure after salvage focal ablation. There exists one other published model by Willigenburg et al. [20] in the setting of salvage focal brachytherapy, but this has not been externally validated.

Future Directions

After FORECAST, further prospective, ideally randomised, studies with longer-term follow-up are required addressing both the diagnosis of radiorecurrent disease, and treatment using salvage focal ablation. These will drive development of an optimised diagnostic and therapeutic paradigm.

It will be useful to evaluate how novel radiological parameters may improve models. Between the models by Peters et al. [7] and Willigenburg et al. [20], radiological variables considered were prostate volume, lesion volume, and stage. However, these do not necessarily quantify the likelihood of radiorecurrent disease. The recently-published Prostate Magnetic Resonance Imaging for Local Recurrence Reporting (PI-RR) guidelines provide a 5-point assessment system for mpMRI post-RT [21]. Although not yet prospectively validated, incorporating this score into models may prove beneficial. For example, higher Prostate Imaging-Reporting and Data System (PI-RADS) score in the primary setting is associated with biochemical failure and metastases by 7 years post-RT [22]. Other radiomic parameters may also be important; our group previously evaluated preoperative mpMRI pharmacokinetic quantitative variables [23]. After adjustment for seven factors, the median interstitial space volume independently predicted failure after salvage focal HIFU. Furthermore, PSMA PET/CT is increasingly used in this population to identify any extra-prostatic disease; maximum standardised uptake value (SUV_{max}) values of any visualised intra-prostatic lesion could also be considered.

Last, on the subject of PSMA PET/CT, data from novel imaging techniques could improve diagnostics and modelling. PSMA PET/CT is increasingly replacing other forms of cross-sectional imaging, including ^{18}F-choline PET/CT and bone scan, the standard at the time of the FORECAST trial [6]. PSMA PET/CT may improve patient selection for salvage focal ablation through greater ability firstly to rule out distant disease, and secondly to rule in local recurrence in conjunction with mpMRI [24,25]. When using both ^{68}Ga-PSMA-11 PET/CT and mpMRI, one study found a positive predictive value of 98% with targeted biopsy [26]. PET/MRI and whole-body MRI are also emerging tools that warrant further study [27,28].

Limitations

Strengths of this study include use of prospective, multicentre UK-wide data with few exclusion criteria. Most importantly however, despite using three sources, our validation cohort is ultimately small, reflecting the paucity of patients undergoing salvage focal therapy. A larger sample size that could predict the calibration slope with ideal precision is likely unobtainable given the relatively few centres globally that perform salvage focal ablation. Our cohort is in fact larger than the majority of published cohorts in both salvage radical prostatectomy and salvage focal therapy [4,5,29]. Second, follow-up was limited to 2 years, whereas the model by Peters et al. [7] was developed with follow-up to 4 years. International collaboration with colleagues from non-UK centres to facilitate further validation and refinement of this model is welcomed, particularly with larger datasets and longer-term follow-up.

Third, registry data are limited by less structured follow-up vs protocol-driven follow-up in FORECAST, though this is arguably more representative of clinical practice.

Last, missing data was another limitation, particularly involving DFS interval (35%) and prostate volume (28%). Missing DFS interval data mainly stemmed from incomplete reporting of when primary treatment was completed. For these patients, the year of primary treatment completion was often available. However, as the DFS interval is a variable measured in months, it was decided to omit calculation of this variable if a specific month was not available. Multiple imputation was instead used to impute these data; an approach that is effective in yielding unbiased results even with large proportions of missingness [30]. We argue this is preferable to the alternative strategy of implementing a rule such as assuming the month of finishing treatment is the mid-point of that year.

In conclusion, for patients who have previously undergone prostate RT, this external validation demonstrates that a previously published risk model can, with reasonable performance, predict if a patient will fail salvage focal ablation by 2 years. Its use should be considered to facilitate appropriate patient selection for salvage focal ablation. Additional external validation in large, non-UK cohorts with longer-term follow-up is needed to further evaluate model performance.

Disclosure of Interests

Taimur T. Shah certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g., employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Alexander Light receives funding from the UK National Institute of Health

Funding

Ethics statement

UK ethics committee approval was received for the FORECAST trial under reference 13/LO/1401.

References

1 Results of the NPCA prospective audit in England and Wales for men diagnosed from 1 National Prostate Cancer Audit 2022.

2 Ma TM, Chu FI, Sandler H et al. Local failure events in prostate cancer treated with radiotherapy: a pooled analysis of 18 randomized trials from the meta-analysis of randomized trials in cancer of the prostate consortium (LEVIATHAN). *Eur Urol* 2022; 82: 487–98

3 Golbari NM, Katz AE. Salvage therapy options for local prostate cancer recurrence after primary radiotherapy: a literature review. *Curr Urol Rep* 2017; 18: 63

4 Valle LF, Lehrer EJ, Markovic D et al. A systematic review and meta-analysis of local salvage therapies after radiotherapy for prostate cancer (MASTER). *Eur Urol* 2021; 80: 280–92

5 Khoo CC, Miah S, Connor MJ et al. A systematic review of salvage focal therapies for localised non-metastatic radiorecurrent prostate cancer. *Transl Androl Urol* 2020; 9: 1535–45

6 Shah TT, Kanthabalan A, Otieno M et al. Magnetic resonance imaging and targeted biopsies compared to transperineal mapping biopsies before focal ablation in localised and metastatic recurrent prostate cancer after radiotherapy. *Eur Urol* 2022; 81: 598–605

7 Peters M, Kanthabalan A, Shah TT et al. Development and internal validation of prediction models for biochemical failure and composite failure after focal salvage high intensity focused ultrasound for local radiorecurrent prostate cancer: presentation of risk scores for individual patient prognoses. *Urol Oncol* 2018; 36: 13.e1–10

8 Shah TT, Peters M, Eldred-Evans D et al. Early-medium-term outcomes of primary focal cryotherapy to treat nonmetastatic clinically significant prostate cancer from a prospective multicentre registry. *Eur Urol* 2019; 76: 98–105

9 Reddy D, Peters M, Shah TT et al. Cancer control outcomes following focal therapy using high-intensity focused ultrasound in 1379 men with nonmetastatic prostate cancer: a multi-institute 15-year experience. *Eur Urol* 2022; 81: 407–13

10 Ahmed HU, El-Shater Bosaily A, Brown LC et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 2017; 389: 815–22

11 Nelson W. Hazard plotting for incomplete failure data. *J Qual Technol* 2018; 1: 27–52

12 Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc* 1996; 91: 473–89

13 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26: 565–74

14 Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001; 125: 279–84

15 Riley RD, Collins GS, Ensor J et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. *Stat Med* 2022; 41: 1280–95

16 Zumsteg ZS, Spratt DE, Romesser PB et al. The natural history and predictors of outcome following biochemical relapse in the dose escalation era for prostate cancer patients undergoing definitive external beam radiotherapy. *Eur Urol* 2015; 67: 1009–16

17 Mitsuzuka K, Arai Y. Metabolic changes in patients with prostate cancer during androgen deprivation therapy. *Int J Urol* 2018; 25: 45–53

18 Sternberg CN, Fizazi K, Saad F et al. Enzalutamide and survival in nonmetastatic, castration-resistant prostate cancer. *N Engl J Med* 2020; 382: 2197–206

19 Kanthabalan A, Arya M, Freeman A et al. Intraprostatic cancer recurrence following radical radiotherapy on transperineal template mapping biopsy: implications for focal ablative salvage therapy. *J Urol* 2020; 204: 950–5

20 Willigenburg T, van Son MJ, van de Pol SMG et al. Development and internal validation of multivariable prediction models for biochemical failure after MRI-guided focal salvage high-dose-rate brachytherapy for radiorecurrent prostate cancer. *Clin Transl Radiat Oncol* 2021; 30: 7–14

21 Panebianco V, Villeirs G, Weinreb JC et al. Prostate magnetic resonance imaging for local recurrence reporting (PI-RR): international consensus - based guidelines on multiparametric magnetic resonance imaging for prostate cancer recurrence after radiation therapy and radical prostatectomy. *Eur Urol Oncol* 2021; 4: 868–76

22 Turchan WT, Kauffmann G, Patel P, Oto A, Liauw SL. PI-RADS score is associated with biochemical control and distant metastasis in men with intermediate-risk and high-risk prostate cancer treated with radiation therapy. *Urol Oncol* 2020; 38: 600.e1–8

23 Rakauskas A, Shah TT, Peters M et al. Can quantitative analysis of multi-parametric MRI independently predict failure of focal salvage HIFU

therapy in men with radio-recurrent prostate cancer? *Urol Oncol* 2021; 39: 830.e1–8

24 **Perera M, Papa N, Roberts M et al.** Gallium-68 prostate-specific membrane antigen positron emission tomography in advanced prostate cancer-updated diagnostic utility, sensitivity, specificity, and distribution of prostate-specific membrane antigen-avid lesions: a systematic review and meta-analysis. *Eur Urol* 2020; 77: 403–17

25 **Light A, Ahmed HU, Shah TT**. The unclear role of PET-CT in localized radiorecurrent prostate cancer. *Nat Rev Urol* 2022; 19: 573–4

26 **Rasing M, van Son M, Moerland M et al.** Value of targeted biopsies and combined PSMA PET/CT and mp-MRI imaging in locally recurrent prostate cancer after primary radiotherapy. *Cancers (Basel)* 2022; 14: 781

27 **Jentjens S, Mai C, Ahmadi Bidakhvidi N et al.** Prospective comparison of simultaneous [^{68}Ga]Ga-PSMA-11 PET/MR versus PET/CT in patients with biochemically recurrent prostate cancer. *Eur Radiol* 2022; 32: 901–11

28 **Adeleke S, Latifoltojar A, Sidhu H et al.** Localising occult prostate cancer metastasis with advanced imaging techniques (LOCATE trial): a prospective cohort, observational diagnostic accuracy trial investigating whole-body magnetic resonance imaging in radio-recurrent prostate cancer. *BMC Med Imaging* 2019; 19: 90

29 **Grubmüller B, Jahrreiss V, Brönimann S et al.** Salvage radical prostatectomy for radio-recurrent prostate cancer: an updated systematic review of oncologic, histopathologic and functional outcomes and predictors of good response. *Curr Oncol* 2021; 28: 2881–92

30 **Madley-Dowd P, Hughes R, Tilling K, Heron J**. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol* 2019; 110: 63–73

Correspondence: Taimur T. Shah, Imperial Prostate, Department of Surgery and Cancer, Hammersmith Hospital, 72 Du Cane Road, London W12 0HS, UK.

e-mail: t.shah@imperial.ac.uk

Abbreviations: ADT, androgen-deprivation therapy; C-index, concordance index; DFS, disease-free survival; FORECAST, FOcal RECurrent Assessment and Salvage Treatment (trial); HEAT, HIFU Evaluation and Assessment of Treatment; HIFU, high-intensity focussed ultrasound; ICE, International Cryotherapy Evaluation; IQR, interquartile range; mpMRI, multiparametric MRI; PET, positron emission tomography; PSMA, prostate-specific membrane antigen; RFS, recurrence-free survival; (EB)RT, (external beam) radiotherapy.

Supporting Information

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

Appendix S1. Calculation of sample size.

Table S1. Patient characteristics, both at diagnosis and at time of enrolment in FORECAST or in the HEAT and ICE registries.

Table S2. Patient characteristics of the original development cohort compared to the external validation cohort described in this study [7].

Table S3. The number of patients who experienced each of the individual outcomes comprising the primary composite failure outcome, together with median time-to-event data, or median follow-up time if the outcome was not reached.