original reports

External Validation of the 2003 Leibovich Prognostic Score in Patients Randomly Assigned to SORCE, an International Phase III Trial of Adjuvant Sorafenib in Renal Cell Cancer

Bhavna Oza, MBBS[1]; Tim Eisen, PhD[2]; Eleni Frangou, MSc (Res)[1]; Grant D. Stewart, MBChB[3]; Axel Bex, MD[4,5]; Alastair W.S. Ritchie, MD[6]; Rick Kaplan, MD[1]; Benjamin Smith, BSc[1]; Ian D. Davis, MBBS, PhD[7,8,9]; Martin R. Stockler, MBBS[9,10]; Laurence Albiges, MD[11]; Bernard Escudier, MD[11]; James Larkin, PhD[12]; Steven Joniau, MD[13]; Barry Hancock, MD[14]; Gregers G. Hermann, DMSc[15]; Joaquim Bellmunt, MD[16]; Mahesh K.B. Parmar, DPhil[1]; Patrick Royston, DSc[1]; and Angela Meade, DPhil[1]

abstract

PURPOSE The 2003 Leibovich score guides prognostication and selection to adjuvant clinical trials for patients with locally advanced renal cell carcinoma (RCC) after nephrectomy. We provide a robust external validation of the 2003 Leibovich score using contemporary data from SORCE, an international, randomized trial of sorafenib after excision of primary RCC.

METHODS Data used to derive the 2003 Leibovich score were compared with contemporary data from SORCE. Discrimination and calibration of the metastasis-free survival outcome were assessed in data from patients with clear-cell RCC, using Cox proportional hazards regression, Kaplan-Meier curves, and calculation of Harrell's c indexes. Secondary analyses involved three important SORCE groups: patients with any non–clear-cell subtype, papillary, and chromophobe carcinomas.

RESULTS Four hundred seven recurrences occurred in 982 patients in the Leibovich cohort and 520 recurrences were recorded in 1,445 patients in the primary SORCE cohort. Clear discrimination between intermediate-risk and high-risk SORCE cohorts was shown; hazard ratio 2.74 (95% CI, 2.29 to 3.28), c-index 0.63 (95% CI, 0.61 to 0.65). A hazard ratio of 0.61 (95% CI, 0.53 to 0.70) confirmed poor calibration of the two cohorts. Discrimination was observed in secondary populations, with c-indexes of 0.64 (95% CI, 0.59 to 0.69) for non–clear-cell RCC, 0.63 (95% CI, 0.56 to 0.69) for papillary RCC, and 0.65 (95% CI, 0.55 to 0.76) for chromophobe RCC.

CONCLUSION The 2003 Leibovich score discriminates between intermediate-risk and high-risk clear-cell and non–clear-cell RCC groups in contemporary data, supporting its use for risk stratification in adjuvant clinical trials. Over time, metastasis-free survival for patients with locally advanced RCC has improved. Contemporary data from adjuvant RCC trials should be used to improve prognostication for patients with RCC.

INTRODUCTION

The Leibovich score,[1] published in 2003, is widely used to guide postnephrectomy prognostication for patients with locally advanced renal cell carcinoma (RCC)[2] and for risk-stratifying patients into adjuvant clinical trials.[2]

Leibovich et al developed the score using retrospective data from patients with clear-cell RCC who underwent radical nephrectomy at the US Mayo Clinic between 1970 and 2000. Five features that were significantly associated with time-to-distant metastases ($P < .001$) comprised the final multivariable model: tumor category (6th TNM 2002), regional lymph-node status, maximum tumor diameter, nuclear grade, and presence of tumor necrosis. For clinical application, risk groups were defined

as low (scores 0-2), intermediate (3-5), and high (6 or higher). Five-year metastasis-free probabilities were reported as 97.1%, 73.8%, and 31.2% respectively.[1]

The SORCE trial (ClinicalTrials.gov identifier: NCT00492258), evaluated the effect of sorafenib after nephrectomy and is one of the largest internationally recruiting randomized controlled trial in patients with locally advanced RCC, to date.[3] In SORCE, and now RAMPART (ClinicalTrials.gov identifier: NCT03288532),[4] the 2003 Leibovich score determines participant eligibility and guides their random allocation to trial arms.

Selection of the 2003 Leibovich score for this purpose is supported by its superior discriminative accuracy on direct comparison with several other prognostic scores.[5] Furthermore, the 2003 Leibovich score is

CONTEXT

Key Objective

The 2003 Leibovich score guides the prognostication and the selection of clear-cell and non–clear-cell patients with locally advanced renal cell carcinoma (RCC) into clinical trials. Its up-to-date validation in contemporary data is necessary to support its continued use. To our knowledge, an evaluation of the 2003 Leibovich score's discrimination between risk groups for non–clear-cell RCCs has not previously been demonstrated.

Knowledge Generated

The 2003 Leibovich score demonstrated discriminative accuracy in contemporary clear-cell and non–clear-cell groups, supporting its use for recruiting and guiding the random assignment of participants to adjuvant RCC trials. Outcomes for patients with RCC have improved over time, rendering the 2003 Leibovich score poorly calibrated to contemporary outcomes.

Relevance

We support the use of the 2003 Leibovich score to risk-stratify patients with RCC suspected of being at intermediate or high risk of relapse.

simple to calculate. All score components are tumor-derived and routinely reported on RCC pathology, negating the need for additional expertise or training. Clinical markers such as patient's performance status are not included in the score, reducing the chance of subjective bias.

An external validation of the 2003 Leibovich score, using data from SORCE participants, was prespecified within the SORCE Protocol (online only). We focused on the intermediate-risk and high-risk patients as they are of specific interest for recruitment to adjuvant clinical trials. SORCE provided a large contemporary data set of individual participant data (IPD) with detailed and long follow-up. Unusually for a validation study, we accessed IPD used to derive the 2003 Leibovich score.[1] By creating closely matched data sets, we were able to compute measures of discrimination and calibration,[6-8] to directly compare the performance of the 2003 Leibovich score in the historical and contemporary cohorts. Accordingly, we provide a high-quality evaluation of the Leibovich score's ability to discriminate between patients at intermediate risk and high risk of relapse.

Although the 2003 Leibovich score is used in clinical trials that recruit patients with non–clear-cell RCC, its ability to stratify risk in this group has not been evaluated. Newer prognostic scores[9-11] have been developed (including some specifically for non–clear-cell subtypes), but none are commonly used in clinical trials, where straightforward application is key. We present the first exploration of the 2003 Leibovich score's discriminative accuracy within important histologic SORCE subpopulations: any non–clear-cell, papillary-only, and chromophobe-only carcinomas.

METHODS

Participants: Leibovich Score Calculation

SORCE participants were recruited from July 2007 to April 2013 from 147 centers in seven countries: United Kingdom, Australia, France, Belgium, the Netherlands, Spain, and Denmark, and followed up until July 2019.[3] Only patients with intermediate (3-5) or high ($\geq$ 6) Leibovich scores were included in SORCE.[1,3] Participants with any histology except pure oncocytoma were eligible. Values for components of the 2003 Leibovich score (Data Supplement, online only) were prospectively collected for each participant on random assignment to SORCE.

The 6th TNM 2002 system was used by Leibovich et al and in SORCE. The same nuclear grading system that selects the worst WHO/International Society of Urological Pathology[12] features at each grade was used in both data sets (Data Supplement). In SORCE, this system pragmatically standardized grading across international trial sites and was used for all histologic subtypes including chromophobe and other non–clear-cell RCCs.

The SORCE trial was approved by national regulatory and ethical committees in each participating country and was conducted in accordance with the principles of Good Clinical Practice, the Declaration of Helsinki, and all applicable regulatory requirements and laws. All participants signed an informed consent form before entry into the study.

Participants and Outcomes: Leibovich and SORCE Populations

Two matched cohorts were analyzed. A derivation cohort derived from the 2003 Leibovich data set included patients with clear-cell RCC only and excluded the low-risk group (Leibovich scores 0-2). A validation cohort was derived from intermediate-risk and high-risk clear-cell RCC participants in SORCE. All patients in the 2003 Leibovich data set underwent radical nephrectomy. Partial or radical nephrectomy was permitted in SORCE, reflecting contemporary surgical practice.

The primary outcome was time to metastasis-free survival (MFS), defined as the interval between nephrectomy and

the date of distant metastases. In the study by Leibovich et al,[1] deaths preceding presumed metastasis were treated as censored observations (C. Lohse, personal communication, April 2020). We defined MFS in the same way for this analysis. We censored time to MFS at 10 years in both cohorts to reflect available follow-up data in SORCE.

Secondary exploratory analyses were conducted in the three SORCE subpopulations: patients with any non–clear-cell histology, papillary-only, and chromophobe-only carcinomas.

Statistical Methods

Model validation was performed adhering to transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines[13] (Data Supplement). The time origin used for both cohorts was date of surgery. A survival analysis allowing for late entry was used, [14] capturing the post hoc nonexposure of a SORCE participant to the risk of an MFS event between surgery and random assignment. In SORCE, 56/1711 (3%) dates of surgery were missing; they were estimated by taking a random selection of 56 values from the distribution of observed intervals between surgery and random assignment.

The performance of the 2003 Leibovich score was assessed using discrimination and calibration.[6,7] Discrimination denotes the ability of a model to distinguish between patients who have and have not experienced an event. Calibration relates to a model's predictive accuracy.

Discrimination was assessed graphically by observing the degree of separation between the Kaplan-Meier curves and by the hazard ratio (HR) between intermediate-risk and high-risk Leibovich risk groups in each cohort. We quantified discrimination according to Harrell's c-index,[8] which denotes the proportion of all usable patient pairs in whom the observed and predicted survival times are concordant. The c-index ranges from 0.5 (performance no better than chance) through to 1 (perfect discrimination).

Calibration measures agreement between predicted and observed outcomes. Good calibration is inferred if Kaplan-Meier curves for risk groups in the derivation and validation cohorts are similar. We quantified calibration through the HR of the indicator variable for the two cohorts (0 = derivation cohort, 1 = validation cohort) separately for the two risk groups (0 = intermediate risk, 1 = high risk). An HR around 1 suggests accurate calibration.

We also analyzed the ungrouped Leibovich scores 3,4, …,11, to compare the HRs between the individual scores and a base category (taken as score = 3). We fitted Cox models separately for the derivation and the validation data sets, with each individual score as the explanatory variable and graphed the results.

Furthermore, we analyzed the ungrouped scores as a single entity to compare the discrimination (c-index) of the Leibovich score with that of 2002 TNM staging.

The secondary (exploratory) analyses were conducted with the three SORCE subpopulations using same procedures as with the primary analysis.

All measures were reported with 95% CIs. *P* values were two-sided.

All analyses were performed in STATA (16.1; StataCorp LLC, College Station, TX).

RESULTS

The 2003 Leibovich data included 479 MFS events in 1671 US-based patients who had radical nephrectomies between 1970 and 2000. The SORCE data had 614 MFS events in 1711 patients enrolled between 2007 and 2013 (Fig 1). The derivation cohort included 407 MFS events in 982 patients with a median follow-up of 7.3 years (interquartile range, 3-10 years), whereas the validation cohort included 520 MFS events in 1,445 patients with a median follow-up of 7.2 years (interquartile range, 6.1-8.4 years; Fig 1).

Table 1 describes the demographic, clinical, and histologic characteristics of patients in the 2003 Leibovich data, the derivation cohort, the SORCE data set, and the validation cohort. The validation cohort included more high-risk than intermediate-risk patients (46% *v* 38%). The validation cohort included 652 (45%) patients who had a laparoscopic nephrectomy and 43 (3%) patients who had a partial nephrectomy, whereas all patients in the derivation cohort underwent radical open nephrectomy (Data Supplement). The median time to MFS in the derivation cohort was 9.2 years, whereas in SORCE, this was not reached within 10 years of follow-up.

Primary Analysis Population: Discrimination and Calibration

Discrimination. Figure 2 presents the results of the validation exercise graphically, showing Kaplan-Meier curves of MFS in the intermediate-risk and high-risk groups for each cohort. Figure 2 shows that discrimination between intermediate-risk and high-risk groups in the derivation cohort is substantial but not entirely maintained in the validation cohort. The c-index in the derivation cohort is 0.67 (95% CI, 0.65 to 0.69) compared with 0.63 (95% CI, 0.61 to 0.65) in the validation cohort (*P* = .01, chi square test).

Discrimination between high-risk and intermediate-risk groups in the derivation cohort, with intermediate-risk as the baseline category, is further indicated by an HR of 3.88 (95% CI, 3.18 to 4.74), compared with 2.74 (95% CI, 2.29 to 3.28) in the validation cohort. Thus, discrimination is maintained in the validation cohort, albeit significantly reduced (*P* = .003, interaction analysis) compared with the derivation cohort.

To assess whether the discrimination of the 2003 Leibovich score degrades over follow-up time, c-indexes at one and
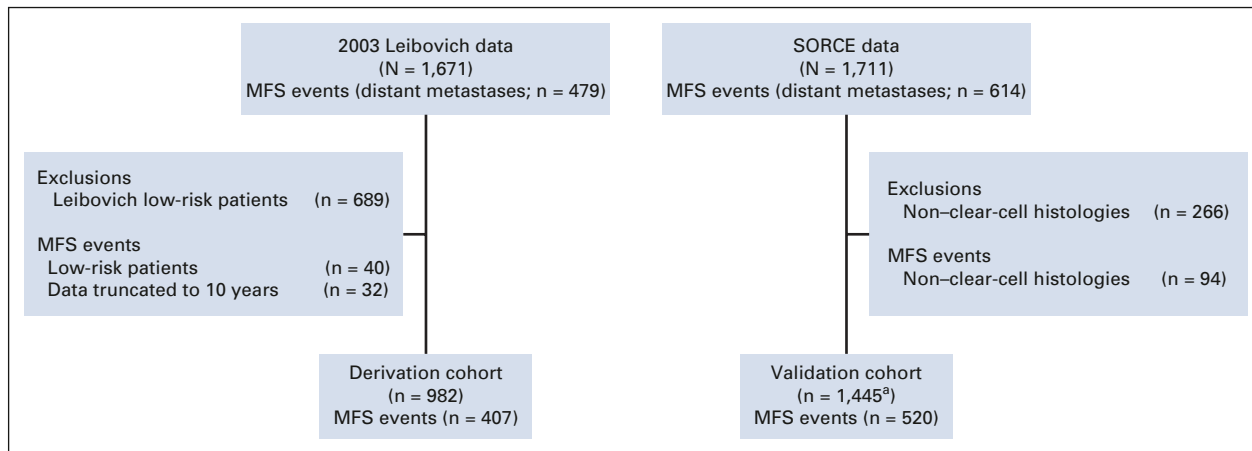
FIG 1. The primary analysis cohorts. MFS: time from nephrectomy to the date of distant metastases; deaths preceding metastasis were censored. [a]A nuclear-grade assignment was missing for one participant, which we imputed singly by substituting the most common nuclear grade value (3), to ensure completeness of the validation data set. MFS, metastasis-free survival.

10 years after nephrectomy were compared in both cohorts (Data Supplement). We show that although c-index values for the Leibovich score reduce over time, the difference is small in both data sets.

Calibration. The validation and derivation survival curves for the intermediate-risk and high-risk groups are not aligned (Fig 2), suggesting poor calibration.

Overall, the MFS rate was 26% lower in the validation than in the derivation cohort (HR = 0.74; 95% CI, 0.65 to 0.85). For the intermediate-risk group, the reduction in MFS rate was 24% (HR = 0.76; 95% CI, 0.61 to 0.94), compared with 46% (HR = 0.54; 95% CI, 0.45 to 0.64) in the high-risk group. The results confirm a distinct lack of calibration between data sets.

Analysis of Ungrouped Leibovich Scores

Figure 3 shows that the HRs comparing individual scores with the reference category (Leibovich score 3) increase markedly as the score increases in both the derivation and validation cohorts, reflecting consistently higher discrimination with increasing Leibovich score. We combined groups with scores of 9 and above because very few patients had score 10 or 11, giving unreliable estimates. See the Data Supplement for Kaplan-Meier curves (Fig 1A), HR values (Data Supplement), and c-indexes for each score (Data Supplement). Lower values of c-index and HR for each score group in the validation cohort confirm in detail that discrimination is maintained, albeit attenuated, in the contemporary cohort.

In both the validation and the derivation cohorts, the collapse of scores 3-5 and 6-11 into two larger prognostic groups (intermediate-risk and high-risk) results in reduced discrimination compared with the original Leibovich score. This is a compromise to achieve a clinically more useful risk stratification tool.

To compare the discrimination of the ungrouped 2003 Leibovich score with that of 2002 TNM, we calculated c-indexes

using the primary analysis data sets. The Leibovich score outperformed the 2002 TNM system in the derivation cohort (c-indexes of 0.72 [SE 0.01] v 0.56 [SE 0.01]) and in the validation cohort (c-indexes of 0.67 [SE 0.01] v 0.56 [SE 0.01]).

Secondary Analyses: Discrimination

Three cohorts were included in the secondary analysis: those with any non–clear-cell RCC (N = 266; MFS events, n = 94), papillary RCC (N = 128; MFS events, n = 49), and chromophobe RCC (N = 96; MFS events, n = 21). Discrimination between intermediate-risk and high-risk groups within each SORCE subcohort was compared with that of the derivation cohort.

Figure 4 shows Kaplan-Meier estimates of MFS in the intermediate-risk and high-risk groups for SORCE non–clear-cell, papillary, and chromophobe populations. The maintained separation between the curves beyond 6 months indicates that the 2003 Leibovich score retains long-term discriminative capability in these SORCE subpopulations. Compared with the derivation cohort (c-index 0.67), we obtained c-indexes of 0.64 (95% CI, 0.59 to 0.69) for the SORCE non–clear-cell cohort, 0.63 (95% CI, 0.56 to 0.69) for SORCE papillary, and 0.65 (95% CI, 0.55 to 0.76) for the SORCE chromophobe group.

An HR of 3.88 (95% CI, 3.18 to 4.74) between risk groups was observed in the derivation cohort, compared with 3.21 (CI, 2.05 to 5.03) for SORCE non–clear-cell patients, 2.61 (95% CI, 1.44 to 4.70) for papillary, and 3.88 (95% CI, 1.56 to 9.61) for the chromophobe cohort. Despite smaller cohort sizes with correspondingly larger imprecision, these results highlight the Leibovich score's preserved discrimination in these SORCE subpopulations.

Secondary Analyses: Calibration

Attenuated calibration between the SORCE subpopulations and the derivation cohort for each risk group is shown by

TABLE 1. Histopathologic Characteristics, Leibovich Score Components, and Median Follow-Up in Leibovich and SORCE Data

Variable at Baseline	2003 Leibovich Data (N = 1,671)	Derivation Cohort (N = 982)	SORCE Data (N = 1711)	Validation Cohort (N = 1,445)
Histologic characteristics, No. (%)				
Histology				
Clear-cell	1,671 (100)	982 (100)	1,445 (84)	1,445 (100)
Papillary	—	—	128 (7)	—
Chromophobe	—	—	96 (6)	—
Collecting duct	—	—	4 (< 1)	—
Other	—	—	38 (2)	—
Other histologies				
Mixed	—	—	8 (21)	—
Sarcomatoid	—	—	23 (61)	—
Unclassified	—	—	5 (13)	—
Translocation	—	—	2 (5)	—
Tumor stage				
pT1a	384 (23)	6 (< 1)	7 (< 1)	5 (1)
pT1b	440 (26)	129 (13)	197 (12)	170 (12)
pT2	335 (20)	335 (34)	400 (23)	298 (10)
pT3a-4	512 (31)	512 (52)	1,107 (65)	972 (67)
Regional lymph node status				
pNx/pN0	1,605 (96)	916 (93)	1,637 (96)	1,405 (97)
pN1/pN2	66 (4)	66 (7)	74 (4)	40 (3)
Tumor size, cm				
< 10	1,312 (79)	623 (63)	1,152 (67)	996 (69)
≥ 10	359 (21)	359 (37)	559 (33)	452 (31)
Nuclear grade[a]				
1	182 (11)	47 (5)	89 (5)	68 (5)
2	786 (47)	284 (29)	440 (26)	374 (26)
3	600 (36)	548 (56)	859 (50)	735 (51)
4	103 (6)	103 (10)	322 (19)	268 (18)
Histologic tumor necrosis[b]				
No	1,232 (74)	561 (57)	774 (45)	671 (46)
Yes	439 (26)	421 (43)	937 (55)	774 (54)
Leibovich score groups in the Leibovich and SORCE data				
Leibovich score group, No. (%)				
Low risk	689 (41)	—	—	—
Intermediate risk	608 (36)	608 (62)	910 (53)	776 (54)
High risk	374 (22)	374 (38)	801 (47)	669 (46)
Median follow-up				
Years (IQR)	7.5 (3.2-10)	7.3 (3.0-10)	7.3 (6.1-8.4)	7.2 (6.1-8.4)

NOTE. Categorical data are presented as No. (%).

Abbreviation: IQR, interquartile range.

[a]The Leibovich and SORCE grading system used simplifies and selects the worst International Society of Urological Pathology features at each grade. For details of grading components, see the Data Supplement.

[b]For the definition of histologic tumor necrosis outlined in SORCE trial protocol, see the Data Supplement.
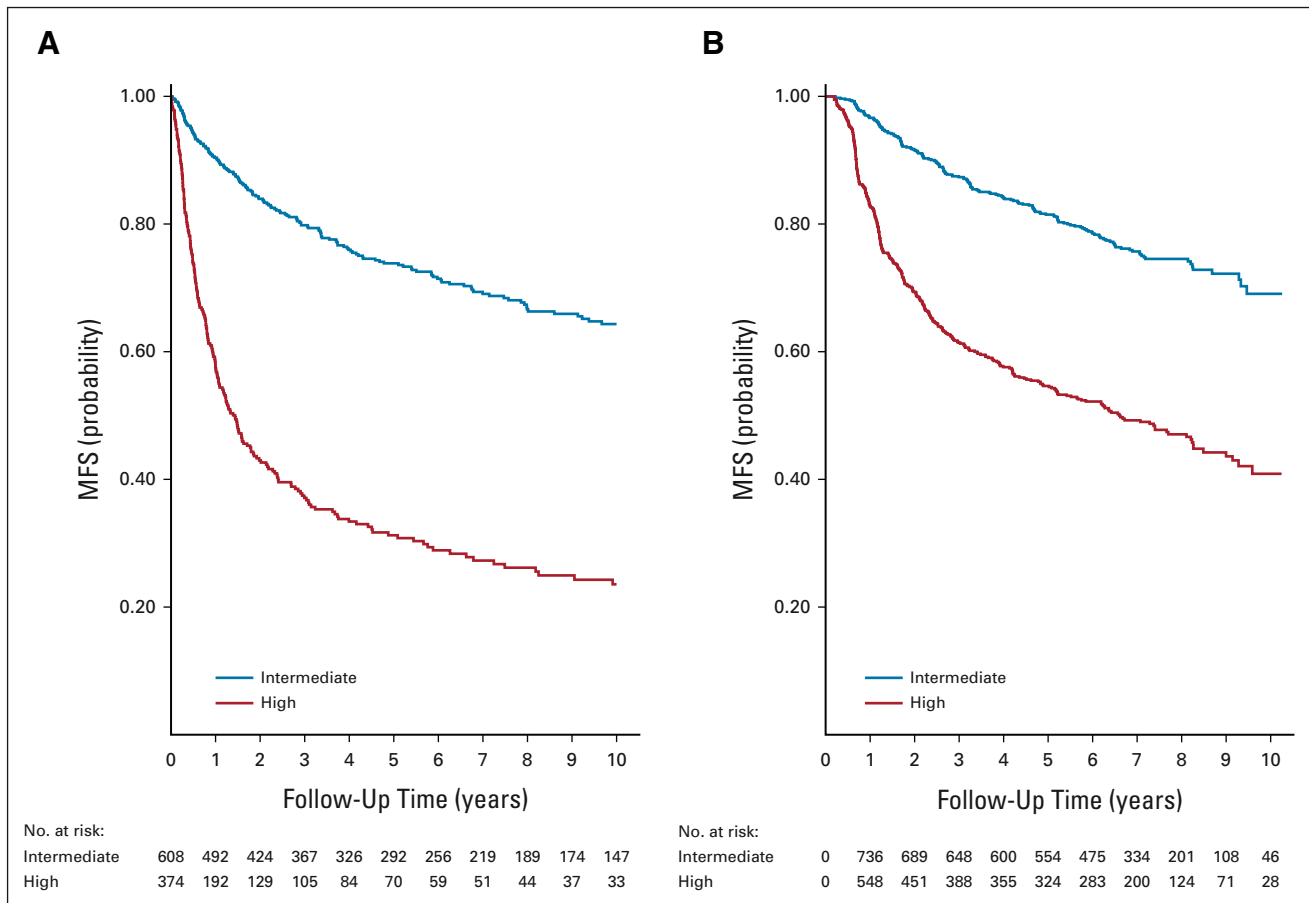
FIG 2. Kaplan-Meier curves for MFS by Leibovich risk group in the (A) derivation and (B) validation cohorts. In the validation cohort Kaplan-Meier plot, the number of patients entering at time 0 is given as 0 in the at-risk tables. It is a consequence of the late entry character of the follow-up data. Patients were not deemed at risk until they were randomly assigned into SORCE, which occurs after t = 0. MFS, metastasis-free survival.

observing the misalignment of the corresponding survival curves (Fig 4). This is quantified by HRs for MFS after fitting a Cox regression model to each risk group separately (Table 2). Five-year relapse probabilities (Table 3) show improved MFS compared with the corresponding derivation cohort in all SORCE subgroups. The difference is most marked between the high-risk groups.

DISCUSSION

Validation of the 2003 Leibovich score using contemporary IPD from a large international trial represents the highest quality of validation, according to the American Joint Committee on Cancer criteria for model selection.[15] We focused on the intermediate-risk and high-risk clear-cell patients, a group commonly recruited to adjuvant clinical trials. This study confirms that the grouped 2003 Leibovich score, although developed two decades ago, largely retains discrimination in the SORCE validation cohort (c-index 0.63; 95% CI, 0.61 to 0.65) when compared with the derivation cohort (c-index 0.67; 95% CI, 0.65 to 0.69). We therefore support its ongoing use for risk stratification in this setting.

Uniquely, we show that the 2003 Leibovich score discriminates comparably between intermediate-risk and high-risk patients in the non–clear-cell SORCE cohort (c-index 0.64; 95% CI, 0.59 to 0.69). Since the non–clear-cell cohort is limited by inherent variability in clinical trajectories, we explored the two largest non–clear-cell subtypes separately: papillary (c-index 0.63; 95% CI, 0.56 to 0.69) and chromophobe groups (c-index 0.65; 95% CI, 0.55 to 0.76). Although the latter analyses are limited by smaller patient numbers, they indicate negligibly attenuated discrimination compared with the derivation cohort.

Some of the immune-oncology–focused adjuvant RCC trials, including IMMOTION010 (ClinicalTrials.gov identifier: NCT03024996) and KEYNOTE-564,[16] use the TNM staging system for patient random assignment to trial arms. We show that discrimination of the 2003 Leibovich score exceeds that of 2002 TNM in the derivation cohort (c-indexes of 0.72 [SE 0.01] v 0.56 [SE 0.01]) and in the validation cohort (c-indexes of 0.67 [SE 0.01] v 0.56 [SE 0.01]). The improvement is noteworthy, considering that a c-index of 0.5 represents a performance that is no better than chance. This finding has implications for using TNM
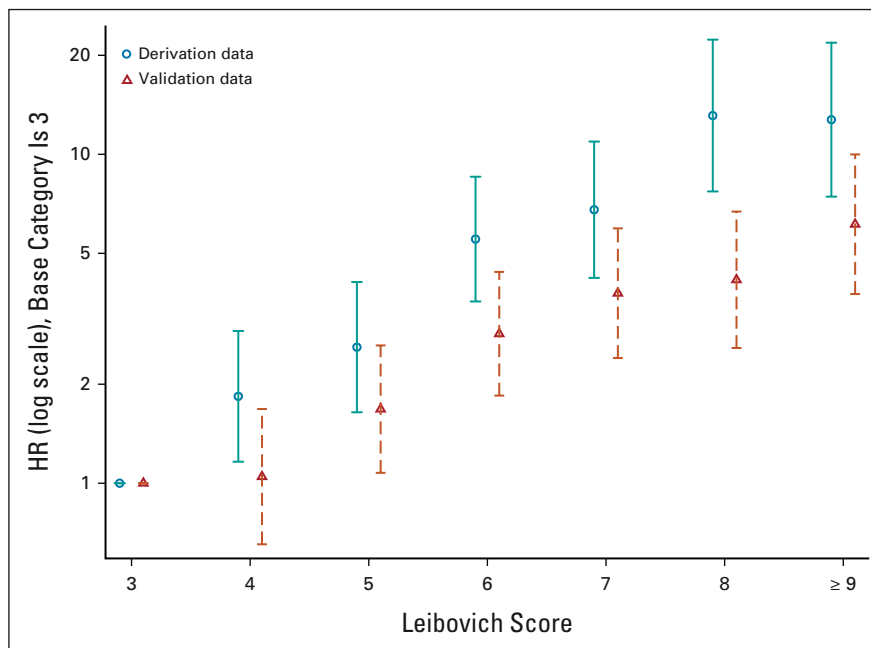
FIG 3. HRs estimated for ungrouped 2003 Leibovich scores in the derivation data set and in the validation data set. Values are presented with 95% CIs. The lowest score (3) in the validation data set is the reference category. HR, hazard ratio.

for participant selection to clinical trials. We also show that the 2003 Leibovich score loses discrimination over follow-up time, with c-indexes of 0.63 at 10 years compared with 0.69 at 1 year after surgery in SORCE (Data Supplement). We suggest that this small difference over long follow-up should not impact on the 2003 Leibovich score's use.

In 2018, Leibovich et al[10] published five scoring systems, modeling progression-free survival (PFS) and cancer-specific survival individually for clear-cell and papillary RCC and PFS for chromophobe carcinomas. A major tradeoff for histologic specificity is added complexity in terms of the number of scoring systems for different subtypes and models that comprise many more components for clear-cell RCC. This is important when considering trial practicalities including standardization and limiting the workload associated with assigning prognostic risk for eligibility purposes. In addition, the 2018 scores offers only minor improvement in discrimination for PFS and cancer-specific survival in clear-cell patients, with internally validated c-indexes of 0.83 and 0.86, respectively, versus 0.82 for MFS for the 2003 score.[1,10]

Overall, the simplicity, practical utility, and maintained discrimination in a multisubtype population shown by the 2003 Leibovich score support its standardized use for risk stratification in adjuvant RCC trials in preference to recently published, yet to be widely externally validated, subtype-specific scores.[9,10]

We were able to perform a robust calibration analysis using IPD from the original Leibovich study, matching risk groups

and unifying the MFS definition across cohorts. We clearly demonstrate longer MFS in patients with intermediate-risk and high-risk clear-cell RCC in the validation cohort (5-year MFS; 78% [CI, 75 to 81] and 52% [CI, 48 to 56], respectively), compared with the corresponding derivation cohorts (5-year MFS; 72% [CI, 67 to 75] and 30% [CI, 25 to 35], respectively). Comparatively longer MFS for contemporary non–clear-cell, papillary, and chromophobe cohorts are also shown (Table 3). On the basis of this, it may be necessary to reconsider trial eligibility for patients with long-term low relapse risk, for example, those with intermediate-risk chromophobe RCC where 5 year MFS approaches 87% (CI, 75-94; Table 3). Overall, better outcomes for patients with locally advanced RCC over time corroborate findings in contemporary literature.[17,18] Improved MFS may be linked to factors such as improved radiologic and pathologic practices over time and the introduction of minimally invasive surgical techniques such as laparoscopic nephrectomy.[19] Differences may additionally reflect an evolution in renal tumor biology over time, driven by changing rates of modifiable risk factors such as obesity and smoking.

Our analysis is not without limitations. First, pathology samples were not centrally reviewed. However, strict guidance for their assessment was provided in the SORCE protocol. Second, patients with low Leibovich risk (score 0-2) were not included this validation, because they are usually cured by surgery or ablation and not usually considered for recruitment to adjuvant trials. We acknowledge that excluding the low-risk group is likely to have resulted in
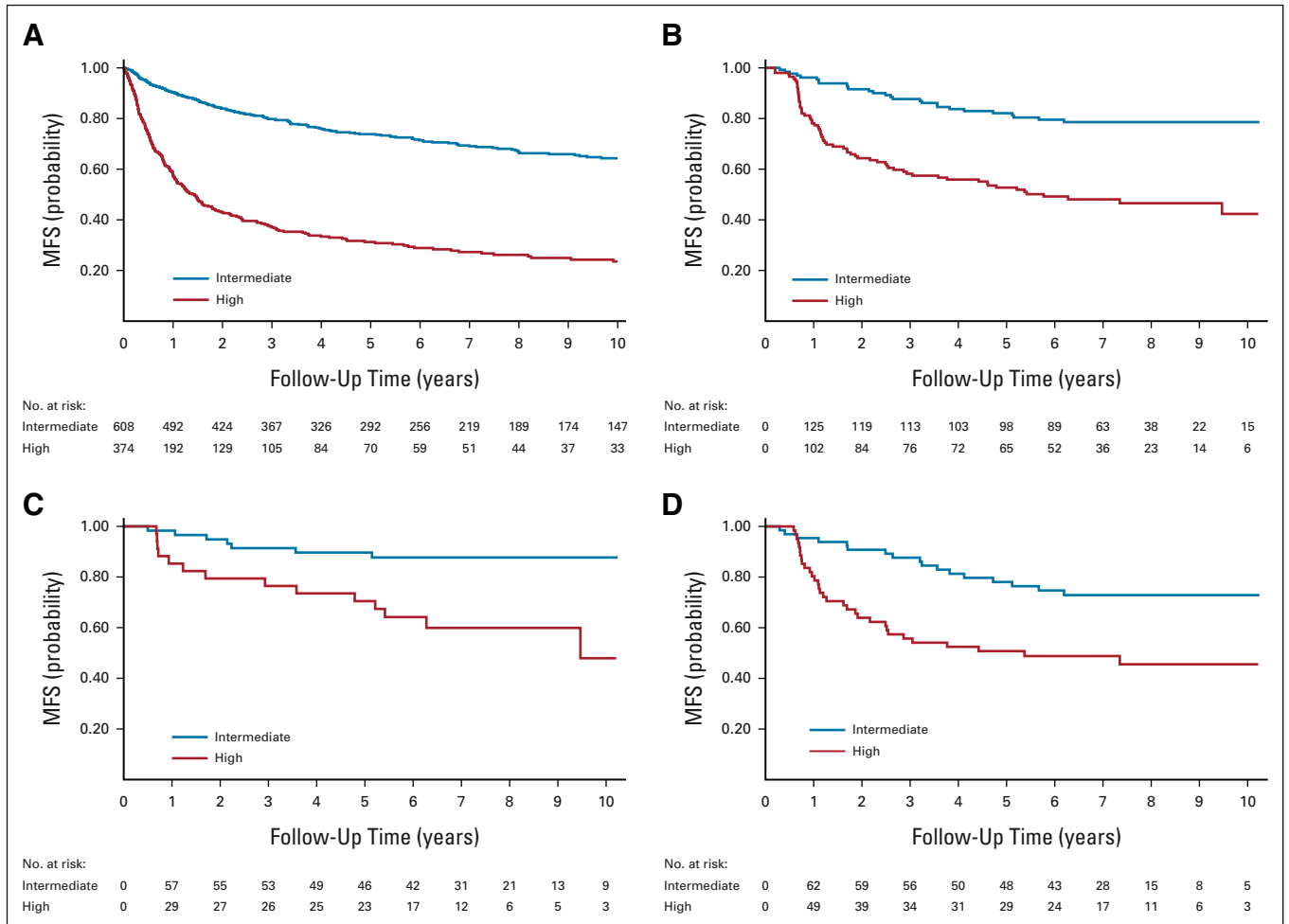
FIG 4. Kaplan-Meier curves for MFS in (A) the derivation cohort and in the SORCE (B) non–clear-cell, (C) chromophobe, and (D) papillary subcohorts stratified by 2003 Leibovich risk group. Derivation cohort included for reference. In the validation cohort Kaplan-Meier plot, the number of patients entering at time 0 is given as 0 in the at-risk tables. It is a consequence of the late entry character of the follow-up data. Patients were not deemed at risk until they were randomly assigned into SORCE, which occurs after t = 0. MFS, metastasis-free survival.

loss of some discrimination compared with that achieved by the complete Leibovich data. Third, our validation was performed using the whole SORCE cohort rather than being restricted to the placebo group. As SORCE showed a clear lack of benefit of sorafenib as an adjuvant strategy after nephrectomy, we considered that including patients from the experimental arms would have no detrimental impact on this analysis.

Finally, patient and tumor characteristics differed between the Leibovich and SORCE cohorts. The median age of SORCE patients was 5 years younger and included higher rates of T3a-4 tumors compared with the Leibovich cohort (67% v 52%). Other differences included higher rates of histologic tumor necrosis in the SORCE cohort (54% v 43%) and more nuclear grade 4 cases (18% v 10%) were present.

TABLE 2. HRs Comparing MFS for Each SORCE Subpopulation to the Derivation Cohort Separately in Intermediate-Risk and High-Risk Groups

Patient Population	SORCE Non–Clear-Cell	95% CI	SORCE Papillary	95% CI	SORCE Chromophobe	95% CI
No. of patients	266		128		96	
No. of MFS events	94		49		21	
Median follow-up, years (IQR)	7.3 (6.1-8.3)		7.3 (6.1-8.4)		7.3 (6.3-8.3)	
Intermediate-risk HRs	0.65	0.43 to 0.97	0.84	0.51 to 1.38	0.36	0.17 to 0.77
High-risk HRs	0.57	0.44 to 0.76	0.59	0.40 to 0.85	0.40	0.23 to 0.69

NOTE. Presented with 95% CI.
Abbreviations: HR, hazard ratio; IQR, interquartile range; MFS, metastasis-free survival.

TABLE 3. Five-Year Survival Probabilities for Metastasis-Free Survival in the Derivation Cohort, the Validation Cohort, and Each SORCE Subcohort

Patient Population	SORCE Validation Cohort (clear-cell), % (95% CI)	SORCE Non–Clear-Cell, % (95% CI)	SORCE Papillary, % (95% CI)	SORCE Chromophobe, % (95% CI)
Derivation, intermediate risk		72 (67 to 75)		
Intermediate risk	78 (75 to 81)	79 (71 to 85)	75 (62 to 84)	87 (75 to 94)
Derivation, high risk		30 (25 to 35)		
High risk	52 (48 to 56)	50 (41 to 58)	49 (36 to 60)	64.9 (45 to 78)

NOTE. Presented with 95% CI.

In time, it may be possible to improve upon outcome prediction in RCC by adapting prognostic scores to include immunologic or genetic biomarkers that show both prognostic and predictive benefit. An example is the transcript-based recurrence score,[20] which adds prognostic information when included with the 2003 Leibovich score. However, as it does not predict response to adjuvant treatment and is expensive and complex, it has not been routinely used.

Alongside prognostic and predictive biomarker studies, a pragmatic step will be to refine the 2003 Leibovich score by further unpicking the characteristics known to drive worse outcomes in RCC. A digital pathology review of SORCE tumor samples is underway. This will allow a comprehensive analysis of the heterogeneity among RCC tumor specimens.[21] It may also reveal further granularity within current 2003 Leibovich score features, to enhance the prognostication and the prediction of recurrence for patients with RCC. A practical goal will be to retain as much of the usefulness and simplicity of the original Leibovich score as possible.

In conclusion, the 2003 Leibovich score is a validated prognostic score which, in contemporary data, discriminates between patients with clear-cell RCC at intermediate risk and high risk of disease recurrence. In addition, it comparably discriminates relapse risk in patients with non–clear-cell, papillary, and chromophobe RCCs in our data set. Over time, MFS rates among patients have improved; therefore, clinicopathologic prognostic scores need to be regularly reviewed. With the wealth of data available from recent RCC trials, there is an opportunity to build upon the 2003 Leibovich score to better reflect the changing landscape of RCC.

AFFILIATIONS
[1]MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, London, United Kingdom
[2]Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
[3]Department of Surgery, University of Cambridge, Cambridge, United Kingdom
[4]Royal Free London NHS Foundation Trust UCL Division of Surgery and Interventional Science, London, United Kingdom
[5]Netherlands Cancer Institute, Amsterdam, the Netherlands
[6]MRC Clinical Trials Unit Trial, London, United Kingdom
[7]Department of Medical Oncology, Eastern Health, Melbourne, Australia
[8]Eastern Health Clinical School, Monash University, Melbourne, Australia
[9]ANZUP Cancer Trials Group, Sydney, Australia
[10]NHMRC Clinical Trials Centre, Medicine, Central Clinical School, Sydney, NSW, Australia
[11]Institut Gustave Roussy, Villejuif, France
[12]The Royal Marsden Hospital, NHS Foundation Trust, London, United Kingdom
[13]Department of Development and Regeneration—Urogenital, Abdominal and Plastic Surgery, Leuven, Belgium
[14]University of Sheffield, Sheffield, United Kingdom
[15]Urology Research Unit, Department of Urology, Herlev, Denmark
[16]Beth Israel Deaconess Medical Center-IMIM Research Lab, Boston, MA

CORRESPONDING AUTHOR
Bhavna Oza, MBBS, MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, 2nd Floor, 90 High Holborn, London WC1V 6LJ, United Kingdom; e-mail: b.oza@ucl.ac.uk.

CLINICAL TRIAL INFORMATION
The SORCE trial (NCT00492258) is closed to recruitment.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST
Disclosures provided by the authors are available with this article at DOI https://doi.org/10.1200/JCO.21.01090.

AUTHOR CONTRIBUTIONS

Conception and design: Tim Eisen, Alastair W.S. Ritchie, Ian D. Davis, Martin R. Stockler, Bernard Escudier, Barry Hancock, Mahesh K.B. Parmar, Patrick Royston, Angela Meade

Administrative support: Benjamin Smith, Angela Meade

Provision of study material or patients: Tim Eisen, Grant D. Stewart, Axel Bex, Ian D. Davis, Martin R. Stockler, Laurence Albiges, Bernard Escudier, James Larkin, Steven Joniau, Barry Hancock, Gregers G. Hermann, Joaquim Bellmunt

Collection and assembly of data: Eleni Frangou, Benjamin Smith, Angela Meade

Data analysis and interpretation: Tim Eisen, Bhavna Oza, Eleni Frangou, Alastair W.S. Ritchie, Rick Kaplan, Ian D. Davis, Martin R. Stockler, Laurence Albiges, Bernard Escudier, James Larkin, Axel Bex, Steven Joniau, Joaquim Bellmunt, Grant D. Stewart, Mahesh K.B. Parmar, Patrick Royston, Angela Meade

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

REFERENCES

1. Leibovich BC, Blute ML, Cheville JC, et al: Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: A stratification tool for prospective clinical trials. Cancer 97:1663-1671, 2003

2. Escudier B, Porta C, Schmidinger M, et al: Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 30: 706-720, 2019

3. Eisen T, Frangou E, Oza B, et al: Adjuvant sorafenib for renal cell carcinoma at intermediate or high risk of relapse: Results from the SORCE randomized phase III intergroup trial. J Clin Oncol 38:4064-4075, 2020

4. Oza B, Frangou E, Smith B, et al: RAMPART: A phase III multi-arm multi-stage trial of adjuvant checkpoint inhibitors in patients with resected primary renal cell carcinoma (RCC) at high or intermediate risk of relapse. Contemp Clin Trials 108:106482, 2021

5. Correa AF, Jegede O, Haas NB, et al: Predicting renal cancer recurrence: Defining limitations of existing prognostic models with prospective trial-based validation. J Clin Oncol 37:2062-2071, 2019

6. Mackillop WJ, Quirt CF: Measuring the accuracy of prognostic judgments in oncology. J Clin Epidemiol 50:21-29, 1997

7. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-387, 1996

8. Harrell FE Jr, Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. JAMA 247:2543-2546, 1982

9. Klatte T, Gallagher KM, Afferi L, et al: The VENUSS prognostic model to predict disease recurrence following surgery for non-metastatic papillary renal cell carcinoma: Development and evaluation using the ASSURE prospective clinical trial cohort. BMC Med 17:182, 2019

10. Leibovich BC, Lohse CM, Cheville JC, et al: Predicting oncologic outcomes in renal cell carcinoma after surgery. Eur Urol 73:772-780, 2018

11. Buti S, Puligandla M, Bersanelli M, et al: Validation of a new prognostic model to easily predict outcome in renal cell carcinoma: The GRANT score applied to the ASSURE trial population. Ann Oncol 28:2747-2753, 2017

12. Amin MB, Epstein JI, Ulbright TM, et al: Best practices recommendations in the application of immunohistochemistry in urologic pathology: Report from the International Society of Urological Pathology consensus conference. Am J Surg Pathol 38:1017-1022, 2014

13. Moons KG, Altman DG, Reitsma JB, et al: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1-W73, 2015

14. Cleves M, Gould WW, Marchenko YV: An Introduction to Survival Analysis Using Stata, Revised 3rd Edition. College Station, TX, Stata Press, 2016

15. Kattan MW, Hess KR, Amin MB, et al: American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. CA Cancer J Clin 66:370-374, 2016

16. Choueiri TK, Tomczak P, Park SH, et al: Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. N Engl J Med 385:683-694, 2021

17. Haas NB, Manola J, Uzzo RG, et al: Adjuvant sunitinib or sorafenib for high-risk, non-metastatic renal-cell carcinoma (ECOG-ACRIN E2805): A double-blind, placebo-controlled, randomised, phase 3 trial. Lancet 387:2008-2016, 2016

18. Vasudev NS, Hutchinson M, Trainor S, et al: UK multicenter prospective evaluation of the Leibovich score in localized renal cell carcinoma: Performance has altered over time. Urology 136:162-168, 2020

19. Ljungberg B, Albiges L, Abu-Ghanem Y, et al: European Association of Urology guidelines on renal cell carcinoma: The 2019 update. Eur Urol 75:799-810, 2019

20. Rini B, Goddard A, Knezevic D, et al: A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: Development and validation studies. Lancet Oncol 16:676-685, 2015

21. Um IH, Scott-Hayward L, Mackenzie M, et al: Computerized image analysis of tumor cell nuclear morphology can improve patient selection for clinical trials in localized clear cell renal cell carcinoma. J Pathol Inform 11:35, 2020

■ ■ ■

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

External Validation of the 2003 Leibovich Prognostic Score in Patients Randomly Assigned to SORCE, an International Phase III Trial of Adjuvant Sorafenib in Renal Cell Cancer

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Tim Eisen
Employment: AstraZeneca, Roche
Leadership: AstraZeneca, Roche
Stock and Other Ownership Interests: AstraZeneca, Roche
Research Funding: Bayer (Inst), Pfizer (Inst), AstraZeneca (Inst)
Travel, Accommodations, Expenses: AstraZeneca, Roche
Other Relationship: Macmillan Cancer Support, Kidney Cancer UK

Grant D. Stewart
Honoraria: Pfizer, Merck, EUSA Pharma
Consulting or Advisory Role: Pfizer, Merck, EUSA Pharma, CMR Surgical
Speakers' Bureau: Pfizer
Research Funding: Pfizer, AstraZeneca, Intuitive Surgical, CRUK
Travel, Accommodations, Expenses: Pfizer

Axel Bex
Consulting or Advisory Role: Pfizer (Inst), Novartis (Inst), Bristol Myers Squibb (Inst), Ipsen (Inst), Eisai (Inst), Genentech (Inst)
Speakers' Bureau: Pfizer, Novartis, Bristol Myers Squibb
Research Funding: Pfizer (Inst)

Rick Kaplan
Research Funding: AstraZeneca (Inst)

Ian D. Davis
Research Funding: Astellas Pharma (Inst), Pfizer (Inst), Roche/Genentech (Inst), MSD Oncology (Inst), AstraZeneca (Inst), Janssen Oncology (Inst), Eisai (Inst), Bayer (Inst), Amgen (Inst), Bristol Myers Squibb (Inst), Movember Foundation (Inst), Exelixis (Inst), Ipsen (Inst), Medivation (Inst), Seattle Genetics (Inst), ANZUP Cancer Trials Group
Patents, Royalties, Other Intellectual Property: International Patent Application No: PCT/US2004/032147 (NY-ESO-1) through Ludwig Institute for Cancer Research

Martin R. Stockler
Research Funding: Astellas Pharma (Inst), Celgene (Inst), Bayer (Inst), Bionomics (Inst), Medivation (Inst), Sanofi (Inst), Pfizer (Inst), AstraZeneca (Inst), Bristol Myers Squibb (Inst), Roche (Inst), Amgen (Inst), Merck Sharp & Dohme (Inst), Tilray (Inst), BeiGene (Inst)
Travel, Accommodations, Expenses: Medivation/Pfizer

Laurence Albiges
Consulting or Advisory Role: Bristol Myers Squibb (Inst), Ipsen (Inst), Roche (Inst), Novartis (Inst), Amgen (Inst), Pfizer (Inst), Astellas Pharma (Inst), Merck (Inst), MSD (Inst), AstraZeneca (Inst), Exelixis (Inst), Janssen (Inst), Eisai (Inst), Corvus Pharmaceuticals (Inst), Peloton Therapeutics (Inst), Bellerophon Theraeutics (Inst)
Research Funding: Bristol Myers Squibb (Inst)
Travel, Accommodations, Expenses: BMS, MSD

Bernard Escudier
Honoraria: Pfizer, Bristol Myers Squibb, Ipsen, Oncorena
Consulting or Advisory Role: Pfizer, Bristol Myers Squibb, Ipsen, AVEO, Oncorena
Research Funding: BMS France (Inst)
Travel, Accommodations, Expenses: Bristol Myers Squibb, Ipsen, MSD

James Larkin
Honoraria: Bristol Myers Squibb, GlaxoSmithKline, Pfizer, Novartis, Roche/Genentech, Incyte, iOnctura, Merck Serono, Eisai, Dynavax Technologies, Cancer Research UK, touchIME, touchEXPERTS
Consulting or Advisory Role: Bristol Myers Squibb, GlaxoSmithKline, Pfizer, Novartis, Boston Biomedical, Incyte, iOnctura, Iovance Biotherapeutics, Immunocore, YKT Corporation, Apple Tree Partners
Research Funding: Pfizer (Inst), Novartis (Inst), MSD (Inst), Bristol Myers Squibb (Inst), Achilles Therapeutics (Inst), Roche (Inst), Nektar (Inst), Covance (Inst), Immunocore (Inst), AVEO (Inst), Pharmacyclics (Inst)
Travel, Accommodations, Expenses: Bristol Myers Squibb, Pfizer, Novartis, Roche/Genentech, AstraZeneca, Incyte, GlaxoSmithKline, Pierre Fabre, Merck Serono, iOnctura, British Uro-Oncology Group (BUG), ESMO, National Cancer Research Institute (NCRI), EUSA Pharma, Syneos Health, Kidney Cancer Association, Bioevents, MedConcept, RV Mais

Steven Joniau
Consulting or Advisory Role: Janssen, AstraZeneca, Bayer, Astellas Pharma
Speakers' Bureau: Astellas Pharma, Janssen, Ipsen
Research Funding: Janssen (Inst), Astellas Pharma (Inst), Ipsen (Inst), Bayer (Inst), Ferring (Inst)
Travel, Accommodations, Expenses: Janssen, Ipsen, Astellas Pharma, Ferring

Joaquim Bellmunt
Stock and Other Ownership Interests: Rainier Therapeutics
Honoraria: UpToDate
Consulting or Advisory Role: Pierre Fabre, Astellas Pharma, Pfizer, Merck, Genentech, Novartis, AstraZeneca/MedImmune, Bristol Myers Squibb
Research Funding: Millennium (Inst), Sanofi (Inst), Pfizer/EMD Serono (Inst)
Travel, Accommodations, Expenses: Pfizer, MSD Oncology, Ipsen

Mahesh K.B. Parmar
Research Funding: AstraZeneca (Inst), Astellas Pharma (Inst), Janssen (Inst), Clovis Oncology (Inst)

Angela Meade
Research Funding: AstraZeneca (Inst)

No other potential conflicts of interest were reported.