New sarculator prognostic nomograms for patients with primary retroperitoneal sarcoma: case volume does matter

Callegaro, D.; Barretta, F.; Raut, C.P.; Johnston, W.; Strauss, D.C.; Honoré, C.; ... ; Gronchi, A.

New Sarculator Prognostic Nomograms for Patients With Primary Retroperitoneal Sarcoma

Case Volume Does Matter

Dario Callegaro, MD,✉ Francesco Barretta, MSc,† Chandrajit P. Raut, MD, MSc,‡ Wendy Johnston, PhD,§ Dirk C. Strauss, MD,‖ Charles Honoré, MD, PhD,¶ Sylvie Bonvalot, MD, PhD,# Mark Fairweather, MD,‡ Piotr Rutkowski, MD, PhD,** Winan J. van Houdt, MD, PhD,†† Rebecca A. Gladdy, MD, PhD,‡‡§§ Fabio Tirotta, MD,‖‖‖ Dimitiri Tzanis, MD,# Jacek Skoczylas, MD,** Rick L. Haas, MD, PhD,¶¶ Rosalba Miceli, PhD,† Carol J. Swallow, MD, PhD,‡‡§§ and Alessandro Gronchi, MD**

Objective: To update the current Sarculator retroperitoneal sarcoma (RPS) prognostic nomograms considering the improvement in patient prognosis and the case volume effect.

Background: Survival of patients with primary RPS has been increasing over time, and the volume–outcome relationship has been well recognized. Nevertheless, the specific impact on prognostic nomograms is unknown.

Methods: All consecutive adult patients with primary localized RPS treated at 8 European and North American sarcoma reference centers between 2010 and 2017 were included. Patients were divided into 2 groups: high-volume centers (HVC, ≥13 cases/year) and low-volume centers (LVC, <13 cases/year). Primary end points were overall survival (OS) and disease-free survival (DFS). Multivariable analyses for OS and DFS were performed. The nomograms were updated by recalibration. Nomograms performance was assessed in terms of discrimination (Harrell C index) and calibration (calibration plot).

Results: The HVC and LVC groups comprised 857 and 244 patients, respectively. The median annual primary RPS case volume (interquartile range) was 24.0 in HVC (15.0–41.3) and 9.0 in LVC (1.8–10.3). Five-year OS was 71.4% (95% CI: 68.3%–74.7%) in the HVC cohort and 63.3% (56.8%–70.5%) in the LVC cohort ($P = 0.012$). Case volume was associated with both OS (LVC vs. HVC hazard ratio 1.40, 95% CI: 1.08–1.82, $P = 0.011$) and DFS (hazard ratio 1.93, 95% CI: 1.57–2.37, $P < 0.001$) at multivariable analyses. When applied to the study cohorts, the Sarculator nomograms showed good discrimination (Harrell C index between 0.68 and 0.73). The recalibrated nomograms showed good calibration in the HVC group, whereas the original nomograms showed good calibration in the LVC group.

Conclusions: New nomograms for patients with primary RPS treated with surgery at high-volume versus low-volume sarcoma reference centers are available in the Sarculator app.

Keywords: liposarcoma, leiomyosarcoma, nomogram, retroperitoneal sarcoma, sarculator, sarcoma

(*Ann Surg* 2024;279:857–865)

Prognosis of patients with primary retroperitoneal sarcoma (RPS) is variable.[1] Tumor-related factors such as histologic type, tumor grade, tumor size, and multifocality; patient-related factors such as patient age; and treatment-related factors such as completeness of surgical resection determine patient survival after surgery.[2] Several nomograms have been developed to predict the individual patient prognosis after resection.[3] Among them, 2 nomograms predicting 7-year overall survival (OS) and 7-year disease-free survival (DFS) have been cited by the AJCC as an alternative option to the TNM staging system and have been included in the Sarculator app.[2]

The 2 Sarculator nomograms were developed on a multi-institutional series of >500 patients who underwent surgery between 1999 and 2009. The nomograms have been externally validated and were found to be reliable and generalizable,[4–7] as they are able both able to separate patients with different prognosis (good discrimination) and that they provide prognostic estimations that are close to what was observed in reality (good calibration).

In a recent collaborative study, we observed that the outcome of patients resected for primary RPS improved significantly over a 15-year period.[8] In particular, 5-year OS

From the *Department of Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; †Department of Biostatistics for Clinical Research, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; ‡Department of Surgery, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; §Department of Surgery, Mount Sinai Hospital, Toronto, ON, Canada; ‖Sarcoma Unit, Department of Academic Surgery, Royal Marsden NHS Foundation Trust, London, United Kingdom; ¶Department of Surgery, Institut Gustave Roussy, Villejuif, France; #Department of Surgery, Institut Curie, Paris, France; **Department of Soft Tissue/Bone Sarcoma and Melanoma, Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland; ††Department of Surgical Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; ‡‡Department of Surgical Oncology, Princess Margaret Cancer Centre, Mount Sinai Hospital, Toronto, ON, Canada; §§Department of Surgery, University of Toronto, Toronto, ON, Canada; ‖‖‖Department of Sarcoma and General Surgery, Midlands Abdominal and Retroperitoneal Sarcoma Unit, University Hospital Birmingham NHS Foundation Trust, Birmingham, United Kingdom; and ¶¶Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands.

✉dario.callegaro@istitutotumori.mi.it.

Presented, in part, at the 2021 Connective Tissue Oncology Society Annual Meeting, Virtual, November 16 to 19, 2021.

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.annalsofsurgery.com.

ISSN: 0003-4932/24/27905-0857

DOI: 10.1097/SLA.0000000000006098

increased of about 10% from 2002 to 2006 (61.2%) to 2012 to 2017 (71.9%) due to a combination of better patient selection, improved quality of surgery, and enhanced perioperative management. As a result, prognostic nomograms, if they are to be used as contemporary tools for patient care, should be adjusted to take into effect the improved survival of patients with RPS.

In addition, population-based and nation-wide studies have highlighted a consistent association between case volume and oncological outcome in patients with RPS, raising the question of whether different prognostic tools should be used in high-volume versus low-volume centers.[9–15]

The aims of our study were (1) to test the hypothesis that the original Sarculator RPS nomograms underestimated prognosis of patients resected in a recent time period and (2) to update the Sarculator RPS nomograms to take into consideration the improved outcome of patients with RPS and the possible case volume effect.

METHODS

All consecutive adult (above 18-year-old) patients with primary (nonrecurrent, nonmetastatic) localized RPS who underwent curative-intent surgery between January 1, 2010, and June 30, 2017, at the following 8 European and North American institutions were included (alphabetic order):

1. Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA.
2. Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.
3. Institut Curie, Paris, France.
4. Institute Gustave Roussy, Villejuif, France.
5. Maria Sklodowska-Curie Institute-Oncology Center, Warsaw, Poland.
6. Mount Sinai Hospital/Princess Margaret Cancer Centre, University of Toronto, Toronto, Canada.
7. Royal Marsden Hospital NHS Foundation Trust, London, UK.
8. The Netherlands Cancer Institute, Amsterdam, the Netherlands.

The study interval (2010–2017) was chosen to avoid overlapping with the series used to develop the original RPS nomograms while maintaining enough follow-up to generate meaningful predictions. Follow-up was updated in May 2021.

To update the nomograms, patients were split into 1 group comprising the higher volume sarcoma reference centers (high-volume sarcoma reference center, HVC, median number of cases/year in the study period $\geq$13) and another group including patients from the lower volume sarcoma reference centers [low-volume sarcoma reference center, low-volume center (LVC), median number of cases/year in the study period <13]. This choice aimed at obtaining different nomogram estimates according to the case volume.

All the participating institutions are part of the Transatlantic Australasian RetroPeritoneal Sarcoma Working Group (TARPSWG). Clinical data were retrieved from prospectively maintained databases in place at each center. The following histologic types were excluded: Ewing sarcoma, alveolar/embryonal rhabdomyosarcoma, desmoid, and desmoplastic small round cell tumor. Adjuvant/neoadjuvant chemotherapy and/or radiotherapy were administered after local multidisciplinary tumor board discussions or as part of clinical trials. Histologic types and subtypes were grouped as follows: well-differentiated liposarcoma, dedifferentiated liposarcoma, leiomyosarcoma, malignant peripheral nerve sheath tumor, solitary fibrous tumor (SFT), undifferentiated pleomorphic sarcoma, and other sarcomas. Tumor grade was determined based upon the Federation Nationale des Centres de Lutte Contre le Cancer criteria throughout the study period in each participating institution. After surgery, follow-up consisted of clinical examination and computed tomography scan of the chest-abdomen-pelvis every 4 to 6 months for the first 5 years and yearly thereafter.

The study was approved by the Institutional Ethics Committee at Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy (Institutional Study Protocol: INT 248/19).

Statistical Methods

Patient cohorts were described using standard descriptive statistics, and the standardized mean difference (SMD) was used as a measure of between-group differences.[16] SMD is considered indicative of a possible between-group imbalance when approaching a value of around 0.3. SMD and the relevance of the observed differences should be assessed from a clinical point of view. The end points were OS and DFS. OS was defined as the time elapsing from surgery to death from any cause. Time was censored at the latest follow-up for patients still alive. DFS was defined as the time elapsing from surgery to relapse or death, whichever occurred first. Time was censored at the latest follow-up for patients still alive and relapse-free. The OS and DFS curves were estimated using the Kaplan–Meier method and compared using the log-rank test. Median follow-up times were estimated by the reverse Kaplan–Meier method from the OS data.[17] Multivariable analyses for OS and DFS were carried out using Cox regression models, including center case volume, patient age, sex, tumor size, tumor grade, histology, multifocality, completeness of resection, intraoperative tumor rupture, and administration of chemotherapy or radiotherapy. The association between 90-day mortality or Clavien-Dindo $\geq$3 complications and center volume were assessed by means of Fisher Exact test. To explore the relative contribution made by 90-day postoperative mortality on OS and DFS, we ran a sensitivity analysis at the 3-month landmark time. In the OS (DFS) sensitivity analysis, the survival time was defined as the time elapsing from the fourth month after surgery to death (Local Recurrence, Distant Metastases) and considering only patients still risk at that specific time-point.

To test the hypothesis that the original Sarculator nomograms underestimated prognosis of patients treated in the study period, their predictive accuracy was assessed in terms of calibration by visual inspection of calibration plots and in terms of discrimination by estimating the Harrell C index in the HVC and LVC cohorts. Where calibration appeared not acceptable/emendable, recalibration was performed, and new calibration was attempted with the new estimated parameters.

In consideration of the median follow-up time and to align our tools with the majority of the prognostic models available in the literature, the time horizon of the predictions was shortened from 7 to 5 years. The prognostic models used to construct the nomograms were already described in a previous study.[2] The differences between original and recalibrated probability estimates were graphically represented. Statistical analyses were carried out with SAS software (SAS Institute, Cary, NC) and R software (version 4.1.2, last access January 13. 2023, 2023).

RESULTS

The HVC and LVC groups included 857 and 244 patients, respectively. The median number of patients per year operated

upon in the 4 HVC was 24 (IQR 15–41.3), while in the 4 LVC was 9.0 (1.8–10.3). Baseline characteristics of the 2 cohorts are summarized in Table 1.

Median follow-up was 75 months (first and third quartile, IQR, 56–97 mo) in the HVC group and 60 months (IQR 48–81) in the LVC group. HVC and LVC groups were overall similar in terms of histology distribution, with more well-differentiated liposarcoma in the LVC cohort (30.7% vs. 23.2%). In terms of tumor grade, there were less G2 tumors in the LVC cohort (23.8% vs. 43.3%) in favor of both G1 and G3 tumors. In the LVC cohort, there were more multifocal tumors (8.6% vs. 3.5%). In terms of treatment-related variables, in the LVC cohort, there were more tumor ruptures (4.5% vs. 2.0%), the resected organ score was lower (2 vs. 3), and chemotherapy was administered less often (8.6% vs. 14.6%).

By the last follow-up, 276 patients in the HVC cohort and 85 patients in the LVC cohort had died, including 52 and 16 patients who died without evidence of disease, respectively. Five-year OS was 71.4% (95% CI 68.3%–74.7%) in the HVC cohort and 63.3% (56.8%–70.5%) in the LVC cohort ($P = 0.012$, Fig. 1A).

In this study, Local Recurrence occurred as the first event in 225 patients in the HVC cohort and 71 patients in the LVC cohort. Distant metastasis occurred as first event in 174 patients in the HVC cohort and 55 patients in the LVC cohort. Five-year DFS was 53.6% (95% CI: 50.2%–57.2%) in the HVC cohort and 39.9% (33.7-47.2%) in the LVC cohort ($P < 0.001$; Fig. 1B).

Ninety-day postoperative mortality was 2.3% in HVC and 5.7% in LVC (Fisher's Exact test $P = 0.011$, Table 1). Ninety-day Clavien-Dindo ≥ 3 postoperative complications did not differ between the 2 groups ($P = 0.246$, Table 1). At the sensitivity analysis at the 3-month landmark time (Supplementary Figure 1, Supplemental Digital Content 1, http://links.lww.com/SLA/E863), OS remained higher in HVC but the difference was no longer significant (P 0.091) while DFS remained significantly higher in HVC than in LVC ($P = 0.002$).

At multivariable analysis (Table 2), case volume significantly predicted OS [LVC vs. HVC hazard ratio (HR) 1.40, 95% CI: 1.08–1.82, $P = 0.011$] and DFS (HR 1.93, 95% CI: 1.57–2.37, $P < 0.001$).

Application of 'Original' RPS Nomograms on the HVC Cohort

To test the hypothesis that the original RPS nomograms did not accurately estimate OS and DFS in the HVC population, the nomograms were tested in the HVC cohort. The calibration plots (Fig. 2A, B) showed an underestimation of patient prognosis. The underestimation mainly impacted higher-risk tumors (tumors with lower predicted OS/DFS). The Harrell C index was 0.73 for OS and 0.68 for DFS.

Recalibration of the RPS Nomograms on the HVC Cohort

Despite the good discrimination as compared with the external validation performed when the nomograms were developed in 2013 (Harrell C index in the original external validation cohort was 0.67 for OS and 0.68 for DFS),[2] a recalibration appeared needed. Since no new patient-related or tumor-related prognostic variables were available and given the good discrimination ability, we proceeded with the recalibration of the nomograms instead of developing new nomograms.

Using the data of the HVC cohort, we estimated two new Cox models for both OS and DFS with the same variables and structures of the Cox models of the original study[2] and calculated their baseline hazard at 60 months from surgery to replace the original baseline hazard. Since the new nomograms were developed modifying the baseline risk, the discriminative ability remained identical to that estimated with the original models. The new, recalibrated nomograms are shown in Figure 3.

TABLE 1. Demographic, Clinical and Pathologic Characteristics of the Higher Volume and Lower Volume Centers Cohorts

	Higher volume centers	Lower volume centers	SMD
	857	244	—
No. patient per year	—	—	—
HVC, median (first and third quartile)	24.0 (15.0; 41.3)	—	—
LVC, median (first and third quartile)	—	9.0 (1.8; 10.3)	—
Sex, n (%)	—	—	0.111
Male	462 (53.9)	118 (48.4)	—
Female	395 (46.1)	126 (51.6)	—
Age	—	—	0.013
Median (first and third quartile)	62.00 (52.0;70.0)	61.00 (52.0; 69.3)	—
Histology	—	—	0.322
WDLPS	199 (23.2)	75 (30.7)	—
DDLPS	383 (44.7)	88 (36.1)	—
LMS	158 (18.4)	41 (16.8)	—
SFT	39 (4.6)	8 (3.3)	—
MPNST	16 (1.9)	12 (4.9)	—
UPS	25 (2.9)	14 (5.7)	—
Other	37 (4.3)	6 (2.5)	—
FNCLCC grade, n (%)	—	—	0.424
I	254 (29.6)	93 (38.1)	—
II	371 (43.3)	58 (23.8)	—
III	232 (27.1)	93 (38.1)	—
Tumor size	—	—	0.054
Median (first and third quartile)	22.0 (15.0; 30.0)	22.0 (12.5; 33.0)	—
Multifocality, n (%)	—	—	0.215
Yes	30 (3.5)	21 (8.6)	—
No	827 (96.5)	223 (91.4)	—
Chemotherapy, n (%)	—	—	0.188
Yes	125 (14.6)	21 (8.6)	—
No	732 (85.4)	223 (91.4)	—
Radiotherapy, n (%)	—	—	0.035
Yes	198 (23.1)	60 (24.6)	—
No	659 (76.9)	184 (75.4)	—
Resected organ score	—	—	0.462
Median (first and third quartile)	3.0 (2.0; 4.0)	2.0 (1.0; 3.0)	—
Tumor rupture, n (%)	—	—	0.143
Yes	17 (2.0)	11 (4.5)	—
No	840 (98.0)	233 (95.5)	—
Completeness of resection, n (%)	—	—	0.174
Macroscopically incomplete (R2)	20 (2.3)	14 (5.7)	—
Macroscopically complete (R0/R1)	837 (97.7)	230 (94.3)	—
90-day Clavien-Dindo ≥3	—	—	0.083
Yes	171 (20.0)	57 (23.4)	—
No	686 (80.0)	187 (76.6)	—
90-day mortality	—	—	0.174
Yes	20 (2.3%)	14 (5.7%)	—
No	837 (97.7%)	230 (94.3%))	—

DDLPS indicates dedifferentiated liposarcoma; FNCLCC, Federation Nationale des Centres de Lutte Contre le Cancer; LMS, leiomyosarcoma; MPNST, malignant peripheral nerve sheath tumor; SFT, solitary fibrous tumor; UPS, undifferentiated pleomorphic sarcoma; WDLPS, well-differentiated liposarcoma.

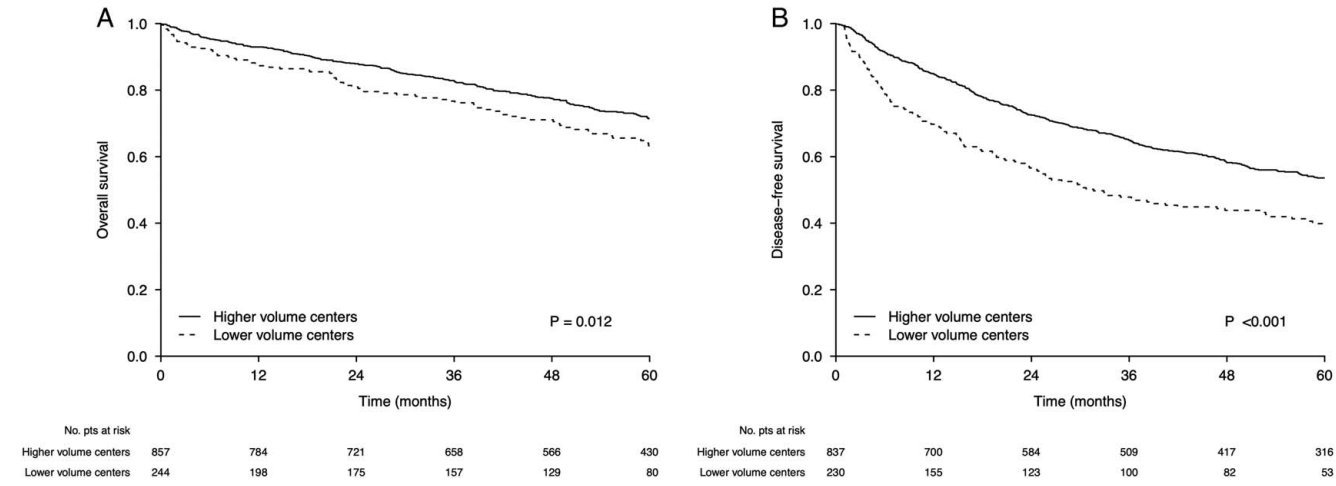

FIGURE 1. Kaplan–Meier curves for overall survival (A) or disease-free survival (B) of higher volume and lower volume centers cohorts.

Internal Validation of the Recalibrated Nomograms on the HVC Cohort

In the calibration plots, when the recalibrated nomograms were tested on the HVC cohort (internal validation, Fig. 2C, D), predicted and observed survival were in good agreement. In particular, the predictions generated by the recalibrated nomograms closely resembled the observed reality, surpassing the predictions generated by the original nomograms. The magnitude of the benefit derived from the recalibration can be visually assessed by comparing the plots of the original nomograms when tested on the HVC cohort (Fig. 2A, B) with the plots of the recalibrated nomograms when tested on the HVC cohort (Fig. 2C, D).

The recalibration of the nomograms resulted, given the same patient and tumor characteristics, in higher survival estimates. A graphical representation of the differences between the original and new probability estimates for both OS and DFS are shown in

TABLE 2. Results From the Multivariable Cox Models on the 2 End Points Analyzed

	OS			DFS		
	HR	**95% CI**	***P***	**HR**	**95% CI**	***P***
Age, y	—	—	< 0.001	—	—	0.002
70 vs 52	1.76	(1.49, 2.07)	—	1.26	(1.10, 1.44)	—
Sex	—	—	< 0.001	—	—	0.003
Male vs female	1.48	(1.19–1.85)	—	1.29	(1.09, 1.54)	—
Tumor size, cm	—	—	0.040	—	—	0.003
30 vs 14	1.24	(1.00, 1.54)	—	1.35	(1.13, 1.60)	—
Histologic type	—	—	0.759	—	—	0.014
WDLPS vs SFT	1.32	(0.50, 3.51)	—	1.69	(0.82, 3.46)	—
DDLPS vs SFT	1.50	(0.70, 3.21)	—	0.98	(0.56, 1.74)	—
LMS vs SFT	1.32	(0.62, 2.82)	—	1.51	(0.87, 2.64)	—
MPNST vs SFT	1.98	(0.78, 4.97)	—	1.26	(0.60, 2.66)	—
UPS vs SFT	1.65	(0.70, 3.90)	—	1.66	(0.86, 3.18)	—
Other vs SFT	1.76	(0.72, 4.32)	—	1.63	(0.83, 3.23)	—
Tumor grade	—	—	< 0.001	—	—	< 0.001
2 vs 1	2.66	(1.17, 6.02)	—	4.65	(2.55, 8.49)	—
3 vs 1	5.90	(2.64, 13.22)	—	7.88	(4.34, 14.32)	—
Multifocality	—	—	0.157	—	—	0.011
Yes vs no	1.35	(0.89, 2.05)	—	1.63	(1.12, 2.36)	—
Completeness of resection	—	—	< 0.001	—	—	—
R0/R1 vs R2	0.33	(0.21, 0.52)	—	—	—	—
Tumor rupture	—	—	0.067	—	—	< 0.001
Yes vs no	1.66	(0.97, 2.84)	—	2.40	(1.47, 3.93)	—
Radiotherapy	—	—	0.302	—	—	0.005
Yes vs no	0.87	(0.67, 1.13)	—	0.74	(0.60, 0.91)	—
Chemotherapy	—	—	< 0.001	—	—	0.001
Yes vs no	1.84	(1.36, 2.49)	—	1.54	(1.20, 1.98)	—
Group	—	—	0.011	—	—	< 0.001
LVC vs HVC	1.40	(1.08, 1.82)	—	1.93	(1.57, 2.37)	—

DDLPS indicates dedifferentiated liposarcoma; FNCLCC, Federation Nationale des Centres de Lutte Contre le Cancer; LMS, leiomyosarcoma; LVC, low-volume centers; MPNST, malignant peripheral nerve sheath tumor; OS, overall survival; SFT, solitary fibrous tumor; UPS, undifferentiated pleomorphic sarcoma; WDLPS, well-differentiated liposarcoma.

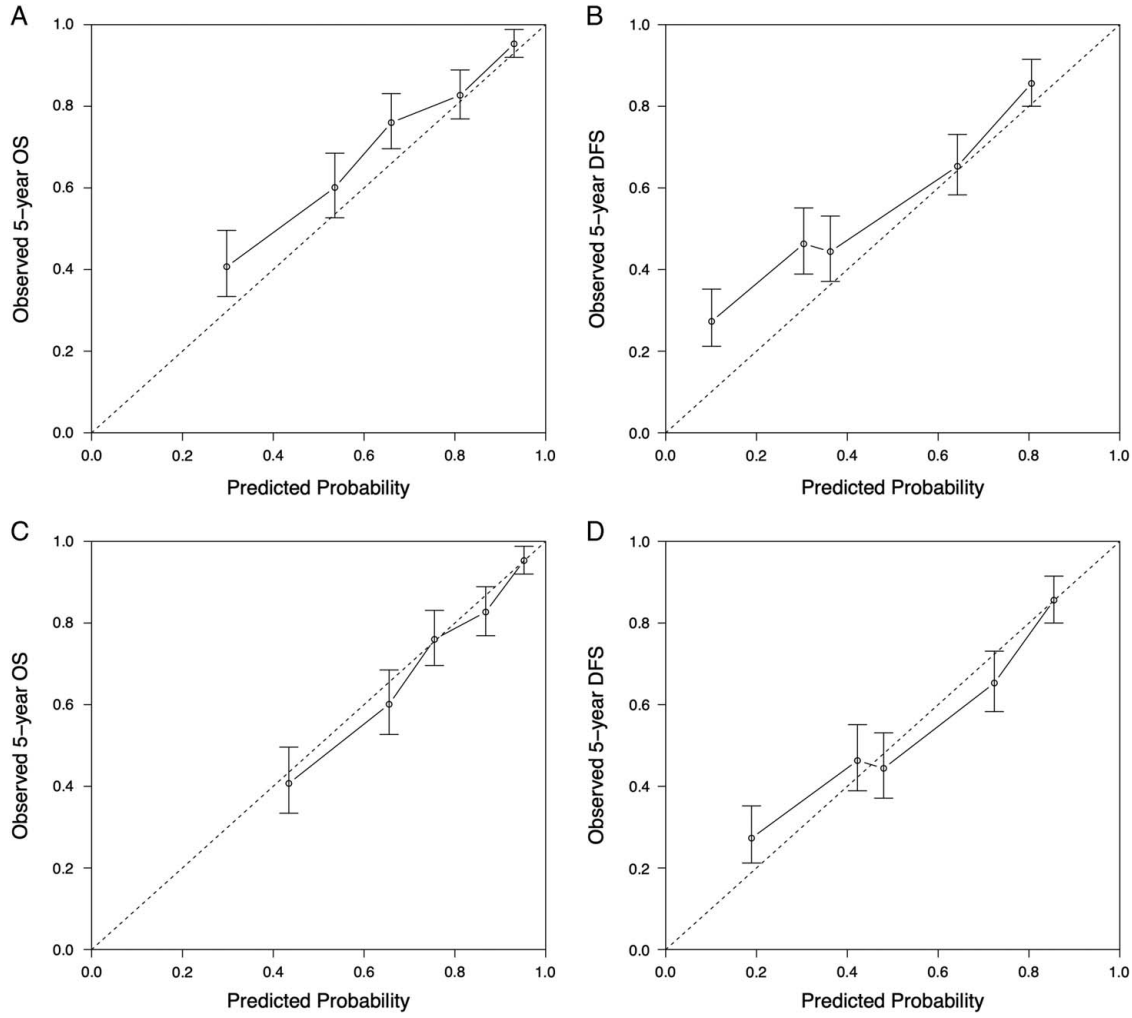

FIGURE 2. Calibration plots for the overall survival (A) and disease-free survival (B) original nomograms when tested on the HVC cohort and calibration plots for the overall survival (C) and disease-free survival (D) recalibrated nomograms when tested on the HVC cohort (internal validation). In these plots, HVC patients were stratified into 5 equally sized subgroups (plots) based on the nomogram-predicted probabilities. Within each subgroup, the average predicted probability (*x*-axis) was plotted against the Kaplan–Meier probability observed in the HVC series (*y*-axis). The vertical lines represent the 95% CIs of the Kaplan–Meier estimates. The dashed line (45-degree line) represents where the plots of a nomogram with a perfect calibration would lie. If the plots fell above the bisector the nomogram would be overly pessimistic, underestimating patient survival. Conversely, if the plots fell below the bisector the nomogram would be overly optimistic, overestimating patient survival. When the original nomograms were tested on the HVC cohort (A, B), most of the plots were positioned above the bisector, indicating that the original nomograms underestimated the prognosis of patients treated at HVC. However, after recalibrating the nomograms (C, D), the plots moved closer to the 45-degree line, indicating that the predictions made with the recalibrated nomograms aligned more closely with the observed reality compared with the predictions made using the original nomograms. OS indicates overall survival.

Figure 4. The maximum differences between original and new probability estimates in OS (+14.0/+14.3%) were observed in patients previously predicted with 22.6% to 38.0% probability. The maximum differences in DFS (+11.5/+11.9%) were observed in patients previously predicted with 22.5% to 40.8% probability.

Application of Original RPS Nomograms on the LVC Cohort

To test the hypothesis that the original RPS nomograms could be used to estimate OS and DFS in the LVC population, the original nomograms were tested in the LVC cohort. The calibration plots (Fig. 5) showed good calibration. This was particularly true for OS prediction plots (all plots lie close to the reference line) while in the DFS nomograms a tendency to overestimation can be seen in medium-low risk patients (medium-high DFS predictions). The Harrell C index was 0.72 for OS and 0.71 for DFS.

DISCUSSION

In this study, first we applied the original Sarculator nomograms for patients with primary RPS to a multicentric cohort of patients with primary RPS who underwent surgery in

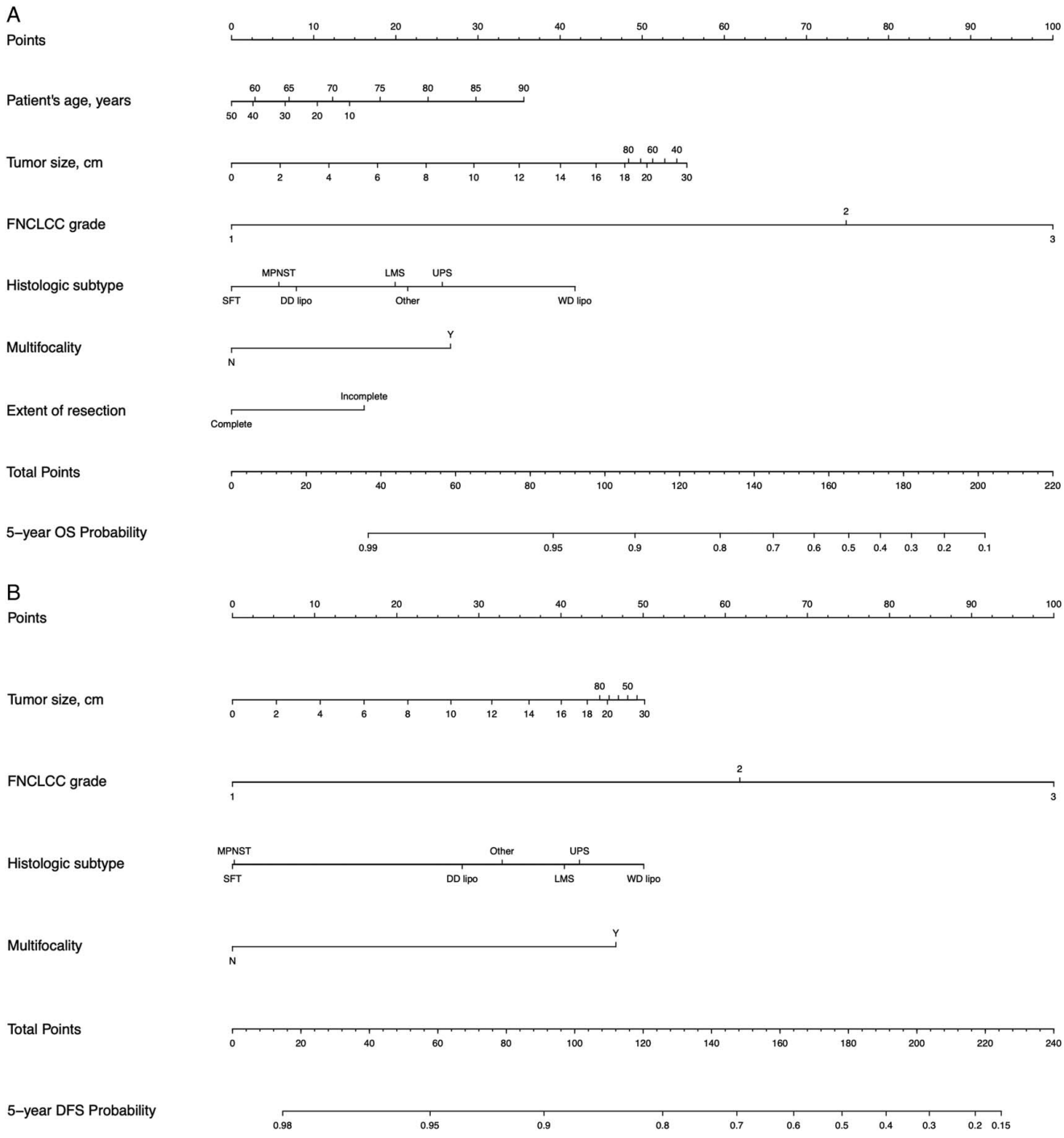

FIGURE 3. Recalibrated nomograms for overall survival (OS, A) and disease-free survival (DFS, B). These nomograms allow to calculate the 5-year overall survival or disease-free survival probability after resection of a primary retroperitoneal sarcoma at a high-volume center. To use the nomograms, the user should calculate the score ("Points" axis) corresponding to each of the prognostic variables (tumor size, FNCLCC grade, histologic subtype, multifocality, and, for the OS nomogram, extent of resection) by drawing vertical lines from each variable axis to the "Points" axis. The scores should be summed, and the total score should be located on the "Total Points" axis. Now, the user should draw a vertical line down to the "5-year OS/DFS Probability" axis to obtain the nomogram-predicted probability. The digital version of the nomograms is available in the free app Sarculator. DD lipo, dedifferentiated liposarcoma; FNCLCC indicates Federation Nationale des Centres de Lutte Contre le Cancer; LMS, leiomyosarcoma; MPNST, malignant peripheral nerve sheath tumor; SFT, solitary fibrous tumor; UPS, undifferentiated pleomorphic sarcoma; WD lipo, well differentiated liposarcoma.

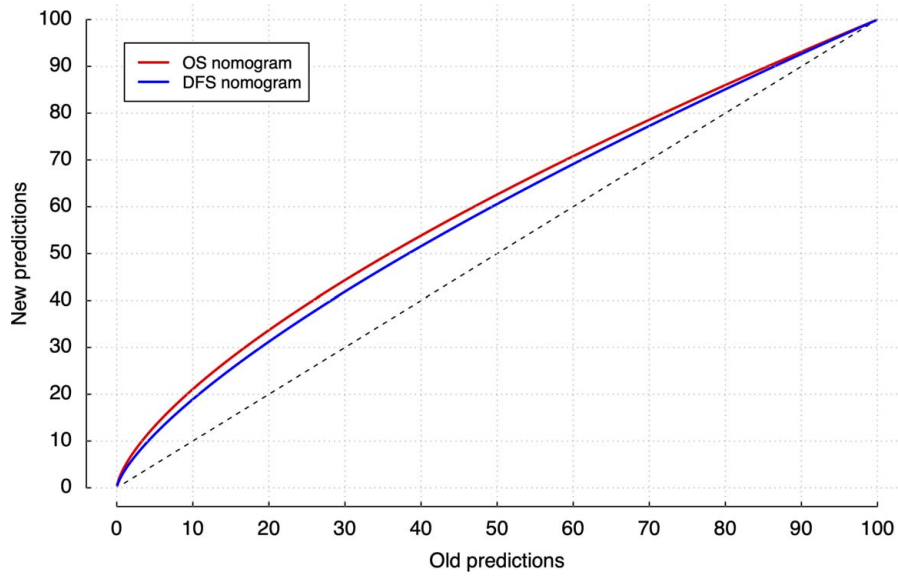

FIGURE 4. Curves of the predicted probabilities of the OS (red line) and DFS (blue line) nomograms after recalibration compared with the predicted probabilities of the original nomograms (dashed line, 45-degree line). In this figure, the distance between the colored lines and the dashed line represents the absolute percentage increase in the survival estimates resulting from the nomograms recalibration. For instance, if the original nomograms estimated a 30% 5-year OS for a particular patient, the recalibrated nomograms would now predict a 5-year OS of ~44% for the same patient. The 14% difference between the old and new predictions can be read from 2 different perspectives. First, as a measure of the survival improvement observed from the period 1999 to 2009 (study period of the original study from Gronchi et al[2]) to the period 2010 to 2017 (current study period) in HVC. Second, since the nomograms for LVC did not need a recalibration, as a measure of the survival difference observed in the current study period (2010-2017) between LVC and HVC.

HVC in a recent time period (2010–2017) and we observed an underestimation of patient prognosis. Then, we recalibrated the OS and DFS Sarculator nomograms by adjusting the nomogram baseline risk to take into effect the improved outcome of patients with RPS treated in HVC. The recalibrated OS and DFS nomograms maintained a good discrimination and compared with the previous nomograms, better predicted prognosis of patients with primary RPS treated in HVC. On the contrary, the original Sarculator nomograms still predicted the outcome of patients treated in LVC with good calibration. As a result, different nomograms are now available for patients treated in high-volume versus low-volume centers.

Nomograms are capable of individualizing prognosis prediction by simultaneously taking into consideration the effect of several prognostic variables. The nomograms prognostic estimates reflect the outcome of the cohort used to develop them.

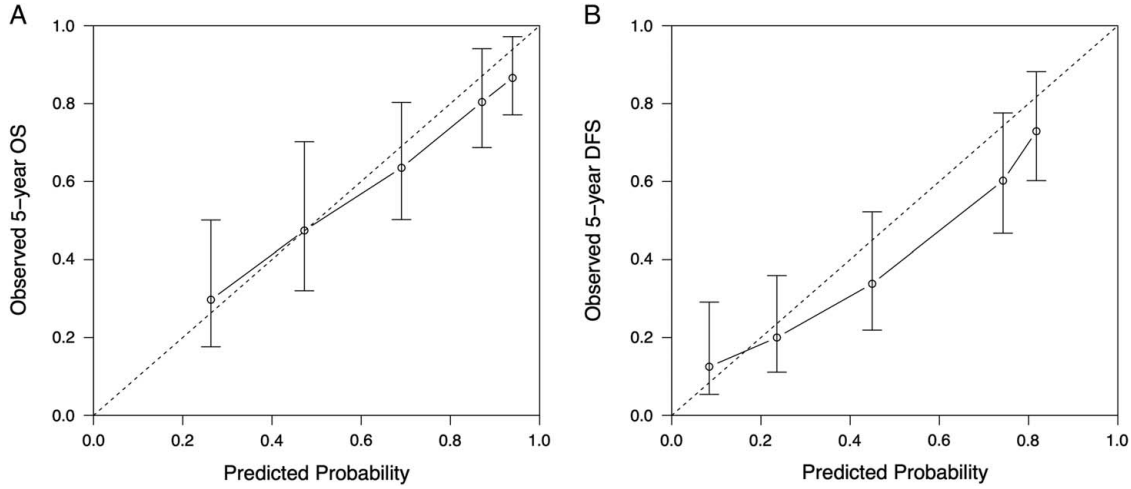

FIGURE 5. Calibration plots for the (A) OS and (B) DFS "original" nomograms when tested on the LVC cohort. OS indicates overall survival.

Any new intervention that affects patient prognosis would make previous nomograms less accurate. Key to this study was that our group recently observed that the outcome of patients with primary RPS treated with surgery in sarcoma reference centers significantly improved over time in the time period 2002 to 2017.[8] This improvement was not related to a single new intervention but rather to a combination of better patient selection, enhanced quality of surgery and improved perioperative management. To address these better outcomes, we sought to determine whether the original Sarculator nomograms, which were developed on a series of patients resected between 1999 and 2009 would underestimate the prognosis of patients treated more recently.

As we have realized an improvement of the prognosis of patients with RPS over time, several key retrospective studies have shown a clear association between RPS treatment at specialized high-volume sarcoma reference centers and improved surgical and oncological outcomes. In particular, treatment at high-volume reference centers has been associated with improved postoperative morbidity and mortality[14,18], better adherence to clinical practice guidelines,[19] and better local progression-free survival and OS.[9,14,20,21] Consistently, in the current study, we observed higher OS and DFS in HVC. Ninety-day postoperative survival contributed to explain, in part, the survival difference. Indeed, at the sensitivity analysis conducted at the 3 months landmark time, the difference in OS was less pronounced while the difference in DFS persisted, although the curves did not separate as much as before (Supplementary Figure 1, Supplemental Digital Content 1, http://links.lww.com/SLA/E863). Since 90-day postoperative complications did not differ between the 2 groups, the difference in early mortality may both reflect different patient selection or different "failure to rescue" between HVC and LVC.

Other studies exploring the volume–outcome relationship in RPS are mainly based upon administrative datasets such as the National Cancer Database, with all the limitations thereof. Specifically, the participating institutions differed not only in terms of case volume but also in the other metrics that might describe the quality criteria needed to identify a center as a sarcoma reference center, such as multidisciplinary, availability of the facilities needed to diagnose and treat patients with RPS and detailed longitudinal follow-up of clinical outcomes.[22]

Our study provides current reference data for standards of care in these rare tumors, where outcomes have realized improvements. First, it was built upon clinic-based prospectively maintained registries with high-quality data. Second, all the participating centers did have in place throughout the study period a dedicated multidisciplinary sarcoma tumor board and the facilities needed to diagnose and treat patients with RPS. Furthermore, they all had ongoing roles in surgical collaboratives being active centers in the Transatlantic Australasian Retroperitoneal Sarcoma Working Group since its inception in 2012. As such, the OS and DFS differences observed between HVC and LVC in the current study (5-y 8% difference in OS and 5-year 14% difference in DFS in favor of HVC) reflect the effect of the case volume purified of all the possible confounders that might affect population-based studies. Given these premises, in the current study, case volume emerged as a main driver of the oncological outcomes. This was also confirmed in the multivariable analyses where the case volume remained significantly associated with OS and DFS after correcting for the main prognostic factors in RPS.

These observations led us to refine our study population by case volume and to create recalibrated nomograms for OS and DFS that could accurately predict the prognosis of patients treated in HVC. In the current study, the median number of patients treated with surgery per year throughout the study period was 24 in the HVC and 9 in LVC. These case volumes were not chosen *a priori* but rather reflect the characteristics of the study population using a cutoff of 13 cases/year to categorize HVC versus LVC, as previously reported in a study merging National Cancer Database data and expert opinions.[10] Nonetheless, they resemble what was observed in a recent population-based analysis of the National Cancer Registration Dataset (England) where high-volume specialist sarcoma centers treated with surgery an average number of 24 patients with primary RPS per year while low-volume specialist sarcoma centers treated an average number of 6 patients with primary RPS per year.[23] In the National Cancer Registration Dataset study, the 11% difference in 5-year OS in favor of high-volume specialist sarcoma centers compare favorably with what was observed in the current study, while in absolute terms, 5-y OS in the current study was higher both in high-volume and low-volume centers (71% vs. 63% in HVC and 63% vs. 52% in LVC).

The choice of recalibrating rather than developing new nomograms is because when the original nomograms were applied to the study cohort, the Harrell C index remained high, around 0.74 for OS and around 0.70 for DFS. This compares favorably with the C index at the original nomograms' validation.[2] These results strengthen the robustness of the Sarculator nomograms and can be viewed as another successful independent external validation. The nomograms remain based upon easy-to-obtain clinical variables readily available after surgery.

After recalibration, the prognostic estimates of the nomograms increased to reflect the improved prognosis of patients operated in a recent time period in high-volume reference centers. This improvement in patient prognosis was evident across the entire probability range, although higher-risk patients benefited the most (Fig. 4). The difference between old and new predictions reflects both the observed improvement in patient prognosis between the period 1999 to 2009 and the period 2010 to 2017 in HVC, as well as the difference observed between HVC and LVC in the time period 2010 to 2017. Since the original nomograms' development cohort (1999–2009) was not stratified by center volume, we cannot determine whether the improvement in patient outcomes over time was solely attributed to HVC or both HVC and LVC.

The main limitations of this study are related to its retrospective nature and to the lack of an external validation of the recalibrated nomograms on an independent series of patients treated in high-volume sarcoma reference centers. Thus, the prospective registry (REtroperitoneal SArcoma Registry, NCT03838718) that is being built by the TARPSWG will represent a unique source of high-quality, contemporary data that might serve to this purpose. Also, the prediction windows of both the recalibrated and the original nomograms were shortened from 7 to 5 years to accommodate the shorter follow-up of the study population. Accordingly, this is a trade-off between choosing a recent cohort while maintaining an adequate follow-up. Finally, the categorization of the centers by volume relies exclusively on primary RPS patients. This is a proxy of a center expertise, and it was associated with oncological outcomes also in previous studies. Nonetheless, we cannot rule out that other, perhaps more robust and explicative, center characteristics such as the number of recurrent cases treated per center, the surgeon volume or expertise of the multidisciplinary team might affect the oncological outcomes but, as for now, we are only able to consider the case volume as a composite metric.

In conclusion, with this study, we updated the Sarculator nomograms for patients with primary RPS to address 2 relevant issues: the improvement of patient prognosis over time and the case volume effect. As a result, there are now 4 nomograms: 2 (OS and DFS) nomograms for patients treated at HVC and 2 nomograms for patients treated at LVC. These nomograms will be implemented in the app Sarculator, and hopefully, the direct appreciation of the individual difference in the predicted prognosis of patients undergoing surgery in HVC versus LVC will be an impetus to increase referral of patients with primary RPS to HVC, as recommended by TARPSWG and by the ESMO guidelines.[22,24]

D.C.: conception and design, acquisition of data, analysis and interpretation of data, article drafting and revision, and final approval of the version to be published. F.B.: conception and design, analysis and interpretation of data, article revision critical for important intellectual content, and final approval of the version to be published. C.P.R.: analysis and interpretation of data, article revision critical for important intellectual content, final approval of the version to be published. W.J.: acquisition of data, article revision critical for important intellectual content, and final approval of the version to be published. D.C.S.: analysis and interpretation of data, article revision critical for important intellectual content and final approval of the version to be published. C.H.: acquisition of data, article revision critical for important intellectual content, and final approval of the version to be published. Sylvie Bonvalot: analysis and interpretation of data, article revision critical for important intellectual content, and final approval of the version to be published. M.F.: acquisition of data, article revision critical for important intellectual content, and final approval of the version to be published. P.R.: analysis and interpretation of data, article revision critical for important intellectual content, and final approval of the version to be published. W.J.v.H.: acquisition of data, article revision critical for important intellectual content, final approval of the version to be published. R.A.G.: acquisition of data, article revision critical for important intellectual content, final approval of the version to be published. F.T.: acquisition of data, article revision critical for important intellectual content and final approval of the version to be published. D.T.: acquisition of data, article revision critical for important intellectual content, and final approval of the version to be published. J.S.: acquisition of data, article revision critical for important intellectual content, and final approval of the version to be published. R.L.H.: analysis and interpretation of data, article revision critical for important intellectual content, final approval of the version to be published. R.M.: conception and design, analysis and interpretation of data, article revision critical for important intellectual content, and final approval of the version to be published. C.J.S.: analysis and interpretation of data, article revision critical for important intellectual content, final approval of the version to be published. A.G.: conception and design, acquisition of data, analysis and interpretation of data, article drafting and revision, and final approval of the version to be published.

REFERENCES

1. Gronchi A, Strauss DC, Miceli R, et al. Variability in patterns of recurrence after resection of primary retroperitoneal sarcoma (RPS): a report on 1007 patients from the Multi-institutional Collaborative RPS Working Group. *Ann Surg*. 2016;263:1002–1009.
2. Gronchi A, Miceli R, Shurell E, et al. Outcome prediction in primary resected retroperitoneal soft tissue sarcoma: histology-specific overall survival and disease-free survival nomograms built on major sarcoma center data sets. *J Clin Oncol*. 2013;31:1649–1655.
3. Callegaro D, Miceli R, Gladdy RA. Prognostic models for RPS patients —attempting to predict patient outcomes. *J Surg Oncol*. 2018;117:69–78.
4. Raut CP, Miceli R, Strauss DC, et al. External validation of a multi-institutional retroperitoneal sarcoma nomogram. *Cancer*. 2016;122:1417–1424.
5. Chou YS, Liu CY, Chang YH, et al. Prognostic factors of primary resected retroperitoneal soft tissue sarcoma: analysis from a single asian tertiary center and external validation of gronchi's nomogram. *J Surg Oncol*. 2016;113:355–360.
6. Squires MH, Ethun CG, Donahue EE, et al. A multi-institutional validation study of prognostic nomograms for retroperitoneal sarcoma. *J Surg Oncol*. 2021;124:829–837.
7. Wong RX, Koh YS, Ong F, et al. Applicability of the Sarculator and MSKCC nomograms to retroperitoneal sarcoma prognostication in an Asian tertiary center. *Asian J Surg*. 2020;43:1078–1085.
8. Callegaro D, Raut CP, Ng D, et al. Has the outcome for patients who undergo resection of primary retroperitoneal sarcoma changed over time? A study of time trends during the past 15 years. *Ann Surg Oncol*. 2021;28:1700–1709.
9. Bonvalot S, Gaignard E, Stoeckle E, et al. Survival benefit of the surgical management of retroperitoneal sarcoma in a reference center: a nationwide study of the French Sarcoma Group from the NetSarc Database. *Ann Surg Oncol*. 2019;26:2286–2293.
10. Villano AM, Zeymo A, Chan KS, et al. Identifying the minimum volume threshold for retroperitoneal soft tissue sarcoma resection: merging national data with Consensus Expert Opinion. *J Am Coll Surg*. 2020;230:151–160.e2.
11. Villano AM, Zeymo A, Chan KS, et al. Variations in retroperitoneal soft tissue sarcoma outcomes by hospital type: a National Cancer Database Analysis. *JCO Oncol Pract*. 2020;16:e991–e1003.
12. Villano AM, Zeymo A, McDermott J, et al. Regionalization of retroperitoneal sarcoma surgery to high-volume hospitals: missed opportunities for outcome improvement. *J Oncol Pract*. 2019;15:e247–e261.
13. Maurice MJ, Yih JM, Ammori JB, et al. Predictors of surgical quality for retroperitoneal sarcoma: Volume matters. *J Surg Oncol*. 2017;116:766–774.
14. Keung EZ, Chiang Y, Cormier JN, et al. Treatment at low-volume hospitals is associated with reduced short-term and long-term outcomes for patients with retroperitoneal sarcoma. *Cancer*. 2018;124:4495–4503.
15. Bagaria SP, Neville M, Gray RJ, et al. The volume-outcome relationship in retroperitoneal soft tissue sarcoma: evidence of improved short- and long-term outcomes at high-volume institutions. *Sarcoma*. 2018;2018:3056562.
16. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34:3661–3679.
17. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials*. 1996;17:343–346.
18. Gutierrez JC, Perez EA, Moffat FL, et al. Should soft tissue sarcomas be treated at high-volume centers? An analysis of 4205 patients. *Ann Surg*. 2007;245:952–958.
19. Heudel PE, Cousin P, Lurkin A, et al. Territorial inequalities in management and conformity to clinical guidelines for sarcoma patients: an exhaustive population-based cohort analysis in the Rhone-Alpes region. *Int J Clin Oncol*. 2014;19:744–752.
20. Toulmonde M, Bonvalot S, Ray-Coquard I, et al. Retroperitoneal sarcomas: patterns of care in advanced stages, prognostic factors and focus on main histological subtypes: a multicenter analysis of the French Sarcoma Group. *Ann Oncol*. 2014;25:730–734.
21. Vos M, Blaauwgeers HGT, Ho VKY, et al. Increased survival of non low-grade and deep-seated soft tissue sarcoma after surgical management in high-volume hospitals: a nationwide study from the Netherlands. *Eur J Cancer*. 2019;110:98–106.
22. Gronchi A, Miah AB, Dei Tos AP, et al. Soft tissue and visceral sarcomas: ESMO-EURACAN-GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2021;32:1348–1365.
23. Tirotta F, Collins S, Bacon A, et al. A comparison of survival outcomes at specialist vs. non-specialist sarcoma centres, for patients with primary retroperitoneal sarcoma: an english population-based coort analysis. Connective Tissue Oncology Society 2021 Annual Meeting.
24. Swallow CJ, Strauss DC, Bonvalot S, et al. Management of primary retroperitoneal sarcoma (RPS) in the adult: an updated consensus approach from the Transatlantic Australasian RPS Working Group. *Ann Surg Oncol*. 2021;28:7873–7888.