CrossMark

ORIGINAL ARTICLE – GASTROINTESTINAL ONCOLOGY

Validation of the MSKCC Gastrointestinal Stromal Tumor Nomogram and Comparison with Other Prognostication Systems: Single-Institution Experience with 289 Patients

Aik-Yong Chok, MBBS, MRCS[1], Brian K. P. Goh, MBBS, MMed, MSc, FRCS[1,2], Ye-Xin Koh, MBBS, MRCS[1], Weng-Kit Lye, MSc[2], John C. Allen Jr., PhD[2], Richard Quek, MBBS, MRCP[3], Melissa C. C. Teo, MBBS, MRCS, FRCS[4], Pierce K. H. Chow, MBBS, FRCS, PhD[1,2], Hock-Soo Ong, MBBS, FRCS[5], Alexander Y. F. Chung, MBBS, FRCS[1], and Wai-Keong Wong, MBBS, FRCS[5]

[1]Department of Hepatopancreatobiliary and Transplantation Surgery, Singapore General Hospital, Singapore, Singapore; [2]Duke-NUS Graduate Medical School, Singapore, Singapore; [3]Division of Medical Oncology, National Cancer Centre, Singapore, Singapore; [4]Division of Surgical Oncology, National Cancer Centre, Singapore, Singapore; [5]Department of Upper Gastrointestinal and Bariatric Surgery, Singapore General Hospital, Singapore, Singapore

ABSTRACT

Purpose. To validate the Memorial Sloan Kettering Cancer Center (MSKCC) prognostic nomogram in a single-institution cohort of patients with gastrointestinal stromal tumors (GISTs), and to compare its predictive accuracy against other established risk classification systems, including the National Institutes of Health (NIH), Armed Forces Institute of Pathology (AFIP), and Joensuu criteria.
Methods. We retrospectively reviewed 289 patients who underwent surgical resection for primary localized GISTs without adjuvant imatinib therapy and compared the actuarial recurrence-free survival (RFS) with the predicted RFS.
Results. Tumors >5 cm in size, with high mitotic index, and which had ruptured were significantly associated with recurrent disease. The 2-year RFS was 77.2 % [95 % confidence interval (CI) 71.6–81.8], and the 5-year RFS was 67.9 % (95 % CI 61.7–73.4). The concordance probability of the nomogram of 2-year RFS was 0.71 (SE 0.02), and 5-year RFS was 0.71 (SE 0.19). The 2-year and 5-year MSKCC nomogram probability calculations and the AFIP criteria gave a better estimation of RFS compared to the NIH ($p < 0.001$) and Joensuu ($p < 0.001$) criteria. There was no significant difference between the predictive accuracy of the nomogram compared to the AFIP criteria.
Conclusions. The MSKCC nomogram slightly underestimated the probability of RFS after surgical resection of GISTs. It was associated with a significantly better predictive accuracy compared to the NIH and Joensuu. This study suggests that there is a wider than expected prognostic divergence between gastric GISTs versus GISTs arising from the small intestine.

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal tumors of the gastrointestinal tract.[1] They may occur anywhere along the alimentary tract, and they may occasionally occur outside the gastrointestinal tract, such as in the omentum and mesentery.[2,3] Significant prognostic heterogeneity has been described with GISTs, which can range from clinically benign to frankly malignant tumors.[4,5] The standard treatment for localized primary GIST is complete surgical resection with clear margins.[6–11] Nonetheless, the risk of recurrence remains even after complete resection.[12,13]

Various classification systems to prognosticate GIST have been proposed. The two most widely accepted risk classification systems are the US National Institutes of Health (NIH) criteria and the Armed Forces Institute of Pathology (AFIP) criteria, which were established in 2001 and 2006, respectively.[4,14] More recently, Joensuu et al. proposed a modified consensus criteria. Table 1

summarizes these three risk classification systems for GISTs. These classification systems have been compared and validated by several investigators.[5,15–18] Several investigators have also proposed modifications to these classification systems, proposing that additional factors such as tumor rupture, tumor ulceration, and mucosal invasion be included in the risk classification.[5,19,21]

Huang et al. recognized the wide prognostic heterogeneity of tumors classified in the NIH high-risk category.[5] They observed that large and mitotically inactive GISTs were less likely to recur compared to mitotically active GISTs, which tended to herald a much poorer outcome. Hence, they proposed that the NIH high-risk category be subdivided into two categories to prognosticate these patients more accurately. Similarly, in 2008, Goh et al. also noted the wide differences in prognostication of GISTs in the AFIP high-risk category and proposed a modification, dividing the high-risk category into a high-risk group and a very high-risk group on the basis of mitotic activity.[19] In another study, Joensuu noted that in addition to tumor size and mitotic count, tumor location and tumor rupture appeared to be important prognostic factors for completely resected GISTs.[20] Hence, these investigators conceptualized a risk classification system for selecting GIST patients for adjuvant therapy based on the Fletcher-NIH consensus criteria as well as taking into account the primary tumor location, as per the AFIP criteria, and the presence of tumor rupture. The importance of tumor rupture as an independent adverse prognostic factor has been highlighted in the European Society of Medical Oncology guidelines for the management of GIST.[21] A tumor, grade, metastasis clinical staging system was proposed by Woodall and colleagues in 2009—a system in the same vein as the tumor, node, metastasis, grade soft tissue sarcoma staging system now used by the American Joint Committee on Cancer.[22]

More recently, in 2009, in a bid to quantify the risk of recurrence of GISTs, the Memorial Sloan Kettering Cancer Center (MSKCC) developed a prognostic nomogram that predicted the risk for tumor recurrence after surgical resection of a localized primary GIST, based on their cohort of 127 patients.[23] The nomogram was then compared with three other existing GIST staging systems using concordance probabilities, and it was found to be slightly more accurate in the prediction of recurrence-free survival (RFS). This nomogram was subsequently validated with two other cohorts of Western patients from the Spanish Group for Research on Sarcomas ($n = 212$) and the Mayo Clinic ($n = 148$).

The present study aimed to validate this prognostic nomogram in a large cohort of Asian patients to determine its applicability in this particular subset of patients. We also compared the predictive accuracy of the GIST nomogram versus current established classification systems.

TABLE 1 Established risk classification systems for GISTs

Risk criteria	Tumor size (cm)	Mitotic count (per 50 HPF)	Primary tumor site
NIH consensus criteria, 2002			
Very low risk	<2	<5	Any
Low risk	≥2 < 5	<5	Any
Intermediate risk	<5	5–10	Any
	5–10	<5	Any
High risk	≥5	≥5	Any
	>10	Any	Any
	Any	>10	Any
AFIP criteria, 2006			
Very low risk	≤2	≤5	Any
	>2 ≤ 5	≤5	Gastric
Low risk	>2 ≤ 5	≤5	Nongastric
	>5 ≤ 10	≤5	Gastric
	≤2	>5	Gastric
Intermediate risk	>10	≤5	Gastric
	>2 ≤ 5	>5	Gastric
	>5 ≤ 10	≤5	Nongastric
High risk	>10	≤5	Nongastric
	≤2	>5	Nongastric
	>2 ≤ 5	>5	Nongastric
	>5 ≤ 10	>5	Any
	>10	>5	Any
Joensuu criteria (modified consensus criteria), 2008			
Very low risk	≤2	≤5	Any
Low risk	>2 ≤ 5	≤5	Any
Intermediate risk	>2 ≤ 5	>5 ≤ 10	Gastric
	≤2	>5 ≤ 10	Any
	>5 ≤ 10	≤5	Gastric
High risk	Any	Any	Tumor rupture
	>10	Any	Any
	Any	>10	Any
	>5	>5	Any
	>2 ≤ 5	>5	Nongastric
	>5 ≤ 10	≤5	Nongastric

GIST gastrointestinal stromal tumor, *HPF* high-power field, *NIH* National Institutes of Health, *AFIP* Armed Forces Institute of Pathology

METHODS

Between 1987 and 2012, a total of 313 patients who underwent complete gross resection of localized primary GIST at Singapore General Hospital were identified from a prospectively maintained surgical database. Of these, 24 patients received adjuvant imatinib therapy and were

excluded from analysis. This study was approved by our institutional review board. Some of these patients have been reported in a previous study.[19] The diagnosis of GIST was confirmed by an expert pathologist using standard pathologic criteria.[19] Mitotic index, which is the number of mitoses per 50 randomly selected microscopic high-power fields (HPF), was calculated. Tumor size was also measured by the pathologist, either before or after formalin fixation. At the time of diagnosis, all the patients showed no evidence of metastatic disease based on staging computed tomography imaging of the abdomen and pelvis. None of these patients was treated with a tyrosine kinase inhibitor at a neoadjuvant or adjuvant setting.

The patients' demographic characteristics and clinicopathologic data were collected. R0 resection was defined as the removal of all gross disease with negative macroscopic and microscopic resection margins on histopathologic examination. The presence of tumor rupture was determined intraoperatively. The postoperative surveillance protocol included physical examination every three months for the first 2 years after surgery, every six months for the next 3 years, and yearly thereafter. Computed tomographic scans of the chest, abdomen, and pelvis were performed every year, and in higher-risk patients every 6 months or earlier if clinically indicated. Endoscopic surveillance was performed yearly. RFS was defined as the time from diagnosis to the time of first documented appearance of the tumor after complete resection on the basis of clinical or radiologic examination.

The MSKCC nomogram (Fig. 1) was applied to our cohort of patients. Points were assigned according to tumor size, number of mitotic figures per 50 HPF, and site of tumor. This was done by drawing a line upward from the corresponding values to the "Points" line. The sum of these three points, plotted on the "Total points" line, corresponded to the nomogram predictions of 2-year and 5-year RFS. The performance of the nomogram was evaluated with the concordance index (C index) and calibration. The C index measured the discriminatory ability of the nomogram, with the interpretation of the C index similar to that of the area under the receiver operating characteristic (ROC) curve.[24] The C index provides the probability that given any two randomly selected patients, the patient who experiences recurrence first has a higher nomogram-predicted probability of recurrence. If both patients experience recurrence at the same time, or if the patient with shorter follow-up does not experience recurrence, then the probability does not apply to those two patients.[23] Second, calibration was assessed by comparing the nomogram-predicted probability of recurrence with the Kaplan–Meier observed RFS for quartiles of patients stratified by nomogram scores after surgical resection.

Statistical analysis was performed by SAS software, version 9.3 (SAS Institute, Cary, NC). Continuous variables are presented as mean (standard deviation) and median (minimum, maximum), and categorical variables are presented as frequency (percentage). The Wilcoxon signed rank test and the χ^2/Fisher's exact test were used to analyze continuous variables and categorical variables, respectively. Two-tailed p values were reported, and a p value of <0.05 was considered to be statistically significant. Survival curve pairwise comparisons within NIH criteria, AFIP criteria, and Joensuu criteria were performed by the log rank test.

RESULTS

The baseline demographic and clinicopathologic characteristics of 289 patients with primary resected GISTs who did not receive adjuvant imatinib therapy are provided in Tables 2 and 3. The median follow-up duration was 61 (range 1–266) months. One hundred fifty-three patients (52.9 %) were male, and the median age of the entire cohort of patients was 61.0 (range 27.0–92.0) years. The median tumor size was 55 (range 3–300) mm, and 170 tumors (58.5 %) had a mitotic index of <5 per 50 HPF.

Univariate analyses (Tables 2, 3) demonstrated that tumor size ≥5 cm (28.0 vs. 3.1 %, relative risk [RR] 8.944, 95 % confidence interval [CI] 3.304–24.214, $p < 0.001$), mitotic count ≥5 per 50 HPF (38.7 vs. 1.8 %, RR 21.905, 95 % CI 6.976–68.779, $p < 0.001$), and tumor rupture (58.8 vs. 14.3 %, RR 4.103, 95 % CI 2.507–6.714, $p < 0.001$) were significantly associated with increased 2- and 5-year recurrence rates. Nongastric location (18.7 vs. 15.9 %, RR 1.173, 95 % CI 0.699–1.968, $p = 0.545$) and positive resection margins (17.4 vs. 16.9 %, RR 1.028, 95 % CI 0.406–2.605, $p = 0.954$) were not associated with tumor recurrence.
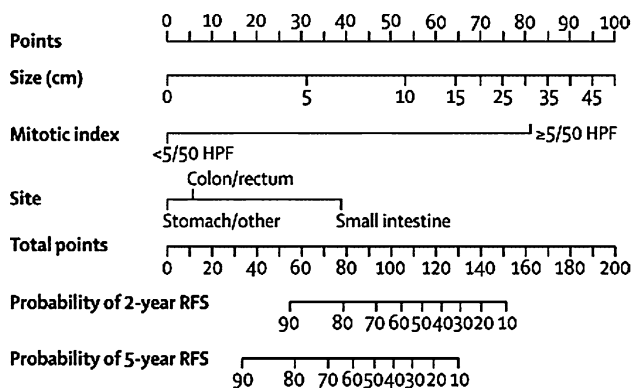

FIG. 1 MSKCC GIST prognostic nomogram to predict probability of 2- and 5-year RFS[25]

TABLE 2 Demographic information and clinicopathologic variables

Variable	Overall (N = 289)	Recurrence event at 2 year (n = 49)	Non recurrence event (n = 240)	OR (95 % CI)[a]	p value[b]
Follow-up period (months)					
Median (range)	61.0 (1.0, 266.0)	40.0 (1.0, 169.0)	67.5 (1.0, 266.0)	1.01 (1.00, 1.02)	0.002
Mean (SD)	74.3 (59.4)	49.5 (37.7)	79.4 (61.8)		
Sex (%)					
Male	153 (52.9)	30 (61.2)	123 (51.3)	Ref.	–
Female	136 (47.1)	19 (38.8)	117 (48.8)	1.37 (0.72, 2.60)	0.338
Age (y)					
Median (range)	61.0 (27.0, 92.0)	58.0 (32.0, 92.0)	62.0 (27.0, 88.0)	1.00 (0.98, 1.03)	0.890
Mean (SD)	60.8 (13.4)	61.5 (15.2)	60.7 (13.0)		
Tumor location (%)					
Stomach	182 (63.0)	29 (59.2)	153 (63.8)	Ref.	–
Duodenum	26 (9.0)	2 (4.1)	24 (10.0)	2.77 (0.61, 12.55)	0.186
Small intestine	53 (18.3)	12 (24.5)	41 (17.1)	0.70 (0.32, 1.51)	0.358
Colon/rectum	20 (6.9)	3 (6.1)	17 (7.1)	1.06 (0.28, 3.97)	0.930
Extraintestinal	8 (2.8)	3 (6.1)	5 (2.1)	0.38 (0.08, 1.75)	0.214
Tumor size (%)					
Median (range), mm	55.0 (3.0, 300.0)	105.0 (7.0, 300.0)	46.0 (3.0, 280.0)	0.98 (0.98, 0.99)	<0.001
Mean (SD), mm	73.6 (61.5)	134.3 (78.8)	61.2 (49.1)		
<50 mm	128 (44.3)	4 (8.2)	124 (51.7)	13.46 (4.64, 39.03)	<0.001
≥50 mm	161 (55.7)	45 (91.8)	116 (48.3)	Ref.	
Mitotic index (%)					
≤5 (/50HPF)	170 (58.8)	3 (6.1)	167 (69.6)	0.03 (0.01, 0.09)	<0.001
>5 (/50HPF)	119 (41.2)	46 (93.9)	73 (30.4)	Ref.	–
Resection margin (%)					
R0	266 (92.0)	45 (91.8)	221 (92.1)	Ref.	–
R1/R2	23 (8.0)	4 (8.2)	19 (7.9)	1.08 (0.34, 3.38)	0.896
Tumor rupture (%)					
Yes	17 (5.9)	10 (20.4)	7 (2.9)	0.10 (0.03, 0.31)	<0.001
No	272 (94.1)	39 (79.6)	233 (97.1)	Ref.	–
NIH criteria (%)					
Very low risk	21 (7.3)	0 (0.0)	21 (8.8)	Inestimable[c]	0.968
Low risk	84 (29.1)	0 (0.0)	84 (35.0)	Inestimable[c]	0.936
Intermediate risk	60 (20.8)	5 (10.2)	55 (22.9)	5.80 (2.12, 15,9)	0.001
High risk	124 (42.9)	44 (89.8)	80 (33.3)	Ref.	–
AFIP criteria (%)					
Very low risk	80 (27.7)	0 (0.0)	80 (33.3)	Inestimable[c]	0.938
Low risk	56 (19.4)	0 (0.0)	56 (23.3)	Inestimable[c]	0.948
Moderate risk	48 (16.6)	7 (14.3)	41 (17.1)	2.69 (1.05, 6.87)	0.039
High risk	105 (36.3)	42 (85.7)	63 (26.3)	Ref.	–
Joensuu criteria (%)					
Very low risk	22 (7.6)	0 (0.0)	22 (9.2)	Inestimable[c]	0.968
Low risk	86 (29.8)	0 (0.0)	86 (35.8)	Inestimable[c]	0.936
Moderate risk	35 (12.1)	3 (6.1)	32 (13.3)	4.70 (1.34, 16.55)	0.016
High risk	146 (50.5)	46 (93.9)	100 (41.7)	Ref.	–

Continuous variables are presented as mean (SD) and categorical variables as n (%)

OR odds ratio, *CI* confidence interval, *NIH* National Institutes of Health, *AFIP* Armed Forces Institute of Pathology

[a] Binary logistic regression model adjusted by follow-up period was used to calculate OR

[b] p value from χ^2 or Fisher's exact test for categorical variables and Wilcoxon–Mann–Whitney test for continuous variables

[c] Logistic regression model had quasicomplete problem, and estimated OR was >999

TABLE 3 Demographic information and clinicopathologic variables

Variable	Recurrence event at 5 years ($n = 63$)	Non recurrence event ($n = 226$)	OR (95 % CI)[a]	p value[b]
Follow-up period (months)				
Median (range)	52.0 (1.0, 177.0)	66.5 (1.0, 266.0)	1.01 (1.00, 1.01)	0.046
Mean (SD)	58.0 (42.4)	78.9 (62.7)		
Sex (%)				
Male	35 (55.6)	118 (52.2)	Ref.	
Female	28 (44.4)	108 (47.8)	1.07 (0.60, 1.89)	0.820
Age (years)				
Median (range)	58.0 (32.0, 92.0)	62.0 (27.0, 88.0)	1.00 (0.98, 1.03)	0.756
Mean (SD)	61.0 (14.7)	60.8 (13.0)		
Tumor location (%)				
Stomach	35 (55.6)	147 (65.0)	Ref.	–
Duodenum	5 (7.9)	21 (9.3)	1.13 (0.40, 3.24)	0.818
Small intestine	15 (23.8)	38 (16.8)	0.63 (0.31, 1.28)	0.199
Colon/rectum	5 (7.9)	15 (6.6)	0.70 (0.23, 2.07)	0.514
Extraintestinal	3 (4.8)	5 (2.2)	0.46 ((0.10, 2.03)	0.302
Tumor size (%)				
Median (range), mm	105.0 (7.0, 300.0)	45.0 (3.0, 280.0)	0.98 (0.98, 0.99)	<0.001
Mean (SD), mm	126.6 (78.0)	58.9 (46.5)		
<50 mm	8 (12.7)	120 (53.1)	8.36 (3.78, 18.52)	<0.001
≥50 mm	55 (87.3)	106 (46.9)	Ref.	–
Mitotic index (%)				
≤5 (/50HPF)	7 (11.1)	163 (72.1)	0.05 (0.02, 0.11)	<0.001
>5 (/50HPF)	56 (88.9)	63 (27.9)	Ref.	–
Resection margin (%)				
R0	59 (93.7)	207 (91.6)	Ref.	–
R1/R2	4 (6.3)	19 (8.4)	1.47 (0.48, 4.53)	0.502
Tumor rupture (%)				
Yes	11 (17.5)	6 (2.7)	0.12 (0.04, 0.36)	<0.001
No	52 (82.5)	220 (97.3)	Ref.	–
NIH criteria (%)				
Very low risk	0 (0.0)	21 (9.3)	Inestimable[c]	0.965
Low risk	1 (1.6)	83 (36.7)	71.33 (9.56, 532,20)	<0.001
Intermediate risk	8 (12.7)	52 (23.0)	4.84 (2.09, 11.19)	<0.001
High risk	54 (85.7)	70 (31.0)	Ref.	–
AFIP criteria (%)				
Very low risk	0 (0.0)	80 (35.4)	Inestimable[c]	0.931
Low risk	2 (3.2)	54 (23.9)	27.47 (6.33, 119,26)	<0.001
Moderate risk	8 (12.7)	40 (17.7)	4.06 (1.69, 9.75)	0.002
High risk	53 (84.1)	52 (23.0)	Ref.	–
Joensuu criteria (%)				
Very low risk	0 (0.0)	22 (9.7)	Inestimable[c]	0.964
Low risk	1 (1.6)	85 (37.6)	60.45 (8.14, 448.86)	<0.001
Moderate risk	5 (7.9)	30 (13.3)	3.70 (1.34, 10.26)	0.012
High risk	57 (90.5)	89 (39.4)	Ref.	–

Continuous variables are presented as mean (SD) and categorical variables as n (%)

OR odds ratio, *CI* confidence interval, *NIH* National Institutes of Health, *AFIP* Armed Forces Institute of Pathology

[a] Binary logistic regression model adjusted by follow-up period was used to calculate OR

[b] p value from χ^2 or Fisher's exact test for categorical variables and Wilcoxon–Mann–Whitney test for continuous variables

[c] Logistic regression model had quasicomplete problem, and estimated OR was >999

The NIH, AFIP, and Joensuu criteria were useful in stratifying patients according to risk of recurrence (Tables 2, 3). Forty-nine patients (17.0 %) were found to have recurrent disease. None of these was classified to have very low-risk or low-risk tumors according to all three risk classification systems (NIH, AFIP, and Joensuu criteria). The majority of tumors with recurrence within 2 years were classified in the high-risk category according to all three risk classification systems.

Figure 2 shows the overall probability of RFS for the entire cohort of patients with GIST for the first 5 years after surgical resection. The 2-year RFS was 77.2 % (95 % CI 71.6–81.8), and the 5-year RFS was 67.9 % (95 % CI 61.7–73.4). Supplementary Fig. 1b–d demonstrates the RFS probability stratified according to the different risk groups within each classification system (NIH, AFIP, and Joensuu). Multiple comparisons between the different risk groups within each criteria were performed. Patients in the high-risk category according to the NIH criteria had significantly higher RFS probability compared to patients in the other risk categories ($p < 0.001$). Similarly, patients with high-risk tumors according to the AFIP and Joensuu criteria were significantly more likely to have tumor recurrence.

In each of the three classification systems (NIH, AFIP, and Joensuu), patients classified in the very low-risk and low-risk categories did not differ significantly in terms of RFS probabilities. The AFIP criteria showed a significant difference in RFS probabilities between intermediate-risk patients and very low-risk patients, while the NIH and Joensuu criteria did not show any difference between these two categories of patients. This is likely because of the relatively small sample sizes of patients in the very low-risk category in the NIH ($n = 21$) and Joensuu ($n = 22$) criteria compared to the AFIP criteria ($n = 80$).
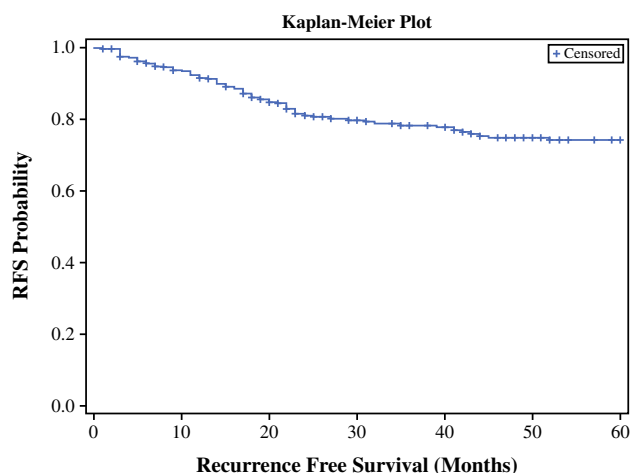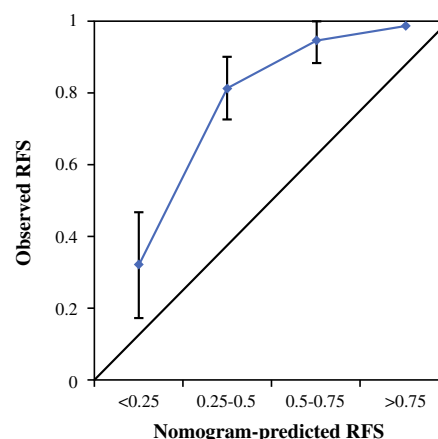

FIG. 3 Calibration of MSKCC nomogram-predicted RFS after surgery during the first 5 years

Figure 3 assesses the calibration of the MSKCC nomogram by plotting the observed RFS against the predicted RFS, where a 45-degree line should be obtained if the predictions are well calibrated. The 2-year nomogram scores of our study cohort were calculated on the basis of the MSKCC nomogram and were divided into four groups by the 25th, 50th, and 75th percentiles. The concordance probability of the nomogram of 2-year RFS was 0.71 (SE 0.02), and 5-year RFS was 0.71 (SE 0.19). Therefore, 71 % of the time, the nomogram correctly predicted the outcome between two randomly selected patients. The Kaplan–Meier plot for observed RFS among our cohort of GIST patients was drawn above the diagonal line, which indicated that the MSKCC nomogram tended to overestimate the probability of recurrence compared to the actual observed recurrence among our patients.

ROC analysis was used to compare the prognostic accuracy of the three GIST risk classification systems (NIH, AFIP, and Joensuu classifications) and the MSKCC nomogram (Fig. 4; Table 4). For the MSKCC nomogram, both the 2-and 5-year predicted probabilities of RFS after surgery for GIST were calculated. The 2- and 5-year nomograms with area under curve (AUC) = 0.87 (95 % CI = 0.82, 0.91) and AUC = 0.87 (95 % CI = 0.83, 0.92), respectively, provided a better estimation than the NIH ($p < 0.001$) and Joensuu ($p < 0.001$) criteria. The 2-year nomogram had a similar ROC curve to the 5-year nomogram, implying that both were similar in terms of predictive ability. There was no significant difference between the performance of the nomogram versus the AFIP criteria ($p = 0.142$). The AUC = 0.85 (95 % CI 0.81–0.88) for the AFIP criteria was also found to be significantly greater than the NIH criteria (AUC = 0.80; 95 % CI 0.76–0.84) ($p = 0.001$) and Joensuu criteria (AUC = 0.77; 95 % CI = 0.73, 0.81) ($p < 0.001$).
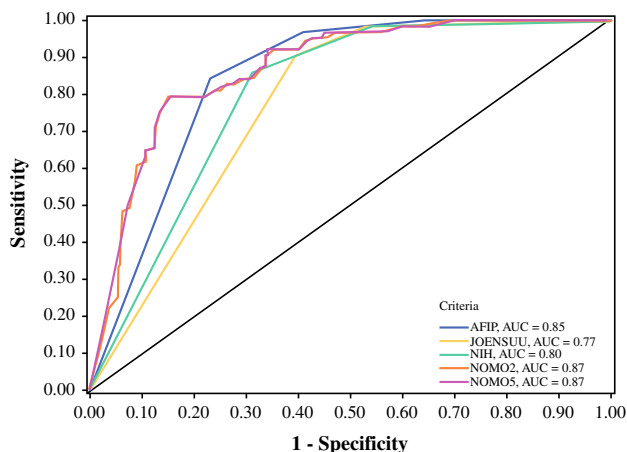

FIG. 2 RFS of GIST during the first 5 years

FIG. 4 ROC curve analysis of risk of GIST recurrence during the first 5 years

TABLE 4 Comparison between receiver operating characteristic curves

Criteria	Difference (95 % CI)	p value
AFIP—NIH (Ref.)	0.05 (0.02, 0.07)	0.001
NOMO2—NIH (Ref.)	0.07 (0.03, 0.11)	<0.001
NOMO5—NIH (Ref.)	0.07 (0.03, 0.11)	<0.001
Joensuu—NIH (Ref.)	−0.03 (−0.05, −0.01)	0.016
AFIP—Joensuu (Ref.)	0.07 (0.04, 0.10)	<0.001
NOMO2—Joensuu (Ref.)	0.10 (0.06, 0.14)	<0.001
NOMO5—Joensuu (Ref.)	0.10 (0.06, 0.14)	<0.001
NOMO2—AFIP (Ref.)	0.02 (−0.01, 0.06)	0.142
NOMO5—AFIP (Ref.)	0.02 (0.01, 0.06)	0.113

CI confidence interval, *AFIP* Armed Forces Institute of Pathology, *NIH* National Institutes of Health, *NOMO* Memorial Sloan Kettering Cancer Center nomogram for 2- and 5-year predicted probabilities

In this study cohort, adjuvant imatinib was routinely available to patients at high risk of recurrence from 2009 after the results from the Intergroup randomized controlled trial demonstrated improved RFS for patients with adjuvant treatment.[25] Before 2009, 223 patients underwent surgical resection, of whom 3 received adjuvant imatinib, whereas from 2009 to 2012, 90 patients underwent resection, of whom 21 received adjuvant treatment. Not surprisingly, the proportion of NIH/AFIP high-risk patients in the cohort resected before 2009 was greater than the cohort resected between 2009 and 2012 (47.7 %/40 % vs. 27.5 %/24.6 %). The C index of the MSKCC 2-year nomogram, MSKCC 5-year nomogram, NIH, AFIP, and Joensuu in the older cohort was 0.71, 0.71, 0.66, 0.71, and 0.66, respectively, which was similar to the overall findings of the present study. However, the C index of the MSKCC 2-year nomogram, MSKCC 5-year nomogram, NIH, AFIP, and Joensuu in the latter cohort from 2009 to 2012 was 0.68, 0.70, 0.73, 0.73, and 0.73, respectively.

DISCUSSION

Accurate prognostic stratification of resected primary localized GISTs is essential to enable clinicians to counsel patients appropriately and select patients more likely to benefit from adjuvant treatment. Risk stratification may also assist the clinician in determining the intensity of postoperative surveillance. The present study validated the MSKCC nomogram in our cohort of patients and demonstrated that it had a superior predictive accuracy compared to the NIH and Joensuu criteria.

In the present study, univariate analyses demonstrated that tumor size, mitotic index, and tumor rupture were significantly associated with tumor recurrence, whereas the association between tumor location and subsequent recurrence was not statistically significant. Interestingly, the two best classification systems to predict actual tumor recurrence in this study (the MSKCC nomogram and the AFIP criteria) used the same three parameters: size, site, and mitotic index of the tumor. With regard to the other two criteria, the NIH consensus criteria only utilized tumor size and mitotic count, and the Joensuu criteria, also known as the modified consensus criteria, took into account one additional variable: the presence of tumor rupture, which deemed a tumor to be high risk regardless of size, site, or mitotic count, or location of tumor. According to all three risk classification systems, a tumor >5 cm in size and >5 mitotic counts per 50 HPF would be regarded as high risk, and a tumor <2 cm in size and <5 mitotic counts per 50 HPF would be regarded as very low risk, irrespective of its location or the presence of tumor rupture.

It is difficult to determine the exact reason why the AFIP criteria and the MSKCC nomogram better predicted RFS than the other 2 risk classification systems when applied to our patient cohort. However, the most likely reason is that both criteria took into account tumor location as an important criterion. The AFIP system drew a wider prognostic divergence between tumors located in the gastric region compared to nongastric tumors. For example, a small tumor (<2 cm in size) with a mitotic index between 5 and 10 per 50 HPF would be classified as intermediate risk by both the NIH and Joensuu criteria, regardless of its location. However, according to the AFIP criteria, if the tumor were located in the stomach, it would be classified as low risk, whereas if it were located outside the stomach, it would be classified as high risk. Similarly, the MSKCC nomogram assigns 0 points to tumors sited in the stomach, 5 points for those in the colon or rectum, and 40 points for small intestine GISTs. It is also important to note that although tumor rupture was found to be significantly associated with an increased risk of tumor recurrence in this study, the low proportion of ruptured tumors (5.9 %) limited its impact when predicting RFS.

When attempting to validate the MSKCC nomogram in this study cohort, according to which approximately half of the patients harbored high-risk tumors, we found that the observed RFS was slightly better than the nomogram-predicted RFS. In this study, the MSKCC nomogram was a better prognostic predictive tool compared to the NIH and Joensuu criteria. There was no statistically significant difference between the predictive ability of the nomogram and the AFIP criteria. These findings concurred with those reported by Gold et al. who found that the MSKCC nomogram was superior to the NIH criteria but not significantly different from the AFIP criteria.[23] The utility of the Joensuu classification system was not analyzed in their study. Similarly, Tanimine et al. from Japan applied the MSKCC nomogram to a small cohort of 60 Asian patients and reported that the nomogram generally overestimated recurrence risk compared to the actuarial RFS.[26] The authors hypothesized that their results could be confounded by the relatively small patient cohort, with a large proportion (>50 %) of very low-risk and low-risk tumors.[26]

The concordance probability of 0.71 obtained in our patient population was similar to that obtained in the three cohorts of patients that were used to construct and validate the MSKCC nomogram.[23] The nomogram predictions of RFS were relatively well calibrated, although the data in the Kaplan–Meier plot above the diagonal line indicated that the MSKCC nomogram tended to overestimate the probability of recurrence. Subset analysis of our patients who underwent resection before and after the introduction of adjuvant imatinib revealed an important limitation of the MSKCC nomogram. The predictive ability of the nomogram was dependent on the proportion of high/low-risk tumors in a particular study cohort. The nomogram tended to overestimate the probability of recurrence, especially for low-risk tumors, and thus its performance tended to be poorer in study cohorts with a high proportion of low-risk tumors.

The provision of imatinib mesylate as an adjuvant treatment has been shown to prolong RFS and overall survival in patients after surgical resection of GISTs.[25,27,28] However, there is no clear consensus regarding the selection of patients for imatinib postoperatively or the duration of adjuvant treatment after surgical resection of primary GIST.[7,21,29] The nomogram provides RFS probabilities on a continuous scale ranging from 10 to 90 %, although it does not define a specific value at which a tumor should be considered high risk or when the provision of adjuvant imatinib is recommended.[30] Therefore, future validation studies of the MSKCC nomogram ought to seek to determine a value, or a range of values, at which a tumor can be considered high risk.

CONCLUSIONS

The MSKCC nomogram and AFIP criteria had the best predictive accuracy for tumor recurrence compared to the NIH and Joensuu risk classification systems in our series of Asian patients. However, it slightly underestimated the probability of RFS after surgical resection of GISTs. Our study also suggests that there is a wider than expected prognostic divergence between gastric GISTs versus GISTs arising from the small intestine.

DISCLOSURE   The authors declare no conflict of interest.

REFERENCES

1. Miettinen M, Lasota J. Gastrointestinal stromal tumors—definition, clinical, histological, immunohistochemical, and molecular genetic features and differential diagnosis. *Virchows Arch.* 2001;438:1–12.
2. Reith JD, Goldblum JR, Lyles RH, Weiss SW. Extragastrointestinal (soft tissue) stromal tumors: an analysis of 48 cases with emphasis on histologic predictors of outcome. *Mod Pathol.* 2000;13:577–85.
3. Goh BK, Chow PK, Kesavan SM, Yap WM, Chung YF, Wong WK. A single-institution experience with eight CD117-positive primary extragastrointestinal stromal tumors: critical appraisal and a comparison with their gastrointestinal counterparts. *J Gastrointest Surg.* 2009;13:1094–8.
4. Miettinen M, Lasota J. Gastrointestinal stromal tumors: pathology and prognosis at different sites. *Semin Diagn Pathol.* 2006;23:70–83.
5. Huang HY, Li CF, Huang WW, et al. A modification of NIH consensus criteria to better distinguish the highly lethal subset of primary localized gastrointestinal stromal tumors: a subdivision of the original high-risk group on the basis of outcome. *Surgery.* 2007;141:748–56.
6. Casali PG, Jost L, Reichardt P, Schlemmer M, Blay JY; ESMO Guidelines Working Group. Gastrointestinal stromal tumours: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol.* 2009;20(Suppl 4):64–7.
7. Demetri GD, von Mehren M, Antonescu CR, et al. NCCN Task Force report: update on the management of patients with gastrointestinal stromal tumors. *J Natl Compr Canc Netw.* 2010; 8(Suppl 2):S1–41.
8. Blackstein ME, Blay JY, Corless C, et al; Canadian Advisory Committee on GIST. Gastrointestinal stromal tumours: consensus statement on diagnosis and treatment. *Can J Gastroenterol.* 2006;20:157–63.
9. Goh BK, Chow PK, Kesavan SM, Yap WM, Chung YF, Wong WK. Outcome after curative resection of large (> or = 10 cm) gastric gastrointestinal stromal tumors: how frequent is adjacent organ involvement and is concomitant distal pancreatectomy necessary? *J Gastrointest Surg.* 2010;14:607–13.
10. Goh BK, Chow PK, Kesavan S, Yap WM, Wong WK. Outcome after surgical treatment of suspected gastrointestinal stromal tumors involving the duodenum: is limited resection appropriate? *J Surg Oncol.* 2008;97:388–91.
11. Chok AY, Koh YX, Ow MY, Allen JC Jr, Goh BK. A systematic review and meta-analysis comparing pancreaticoduodenectomy versus limited resection for duodenal gastrointestinal stromal tumors. *Ann Surg Oncol.* 2014;21:3429–38.

12. DeMatteo RP, Lewis JJ, Leung D, Mudan SS, Woodruff JM, Brennan MF. Two hundred gastrointestinal stromal tumors: recurrence patterns and prognostic factors for survival. *Ann Surg.* 2000;231:51–8.

13. Koh YX, Chok AY, Zheng HL, et al. A systematic review and meta-analysis comparing laparoscopic versus open gastric resections for gastrointestinal stromal tumors of the stomach. *Ann Surg Oncol.* 2013;20:3549–60.

14. Fletcher CD, Berman JJ, Corless C, et al. Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Hum Pathol.* 2002;33:459–65.

15. Rutkowski P, Nowecki ZI, Michej W, et al. Risk criteria and prognostic factors for predicting recurrences after resection of primary gastrointestinal stromal tumor. *Ann Surg Oncol.* 2007;14:2018–27.

16. Nilsson B, Bümming P, Meis-Kindblom JM, et al. Gastrointestinal stromal tumors: the incidence, prevalence, clinical course, and prognostication in the preimatinib mesylate era—a population-based study in western Sweden. *Cancer.* 2005;103:821–9.

17. Ahmed I, Welch NT, Parsons SL. Gastrointestinal stromal tumours (GIST)—17 years' experience from Mid Trent Region (United Kingdom). *Eur J Surg Oncol.* 2008;34:445–9.

18. Hassan I, You YN, Shyyan R, et al. Surgically managed gastrointestinal stromal tumors: a comparative and prognostic analysis. *Ann Surg Oncol.* 2008;15:52–9.

19. Goh BK, Chow PK, Yap WM, et al. Which is the optimal risk stratification system for surgically treated localized primary GIST? Comparison of three contemporary prognostic criteria in 171 tumors and a proposal for a modified Armed Forces Institute of Pathology risk criteria. *Ann Surg Oncol.* 2008;15:2153–63.

20. Casali PG, Blay JY; ESMO/CONTICANET/EUROBONET Consensus Panel of Experts. Gastrointestinal stromal tumours: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2010;21(Suppl 5):v98–102.

21. Joensuu H. Risk stratification of patients diagnosed with gastrointestinal stromal tumor. *Hum Pathol.* 2008;39:1411–9.

22. Woodall CE 3rd, Brock GN, Fan J, et al. An evaluation of 2537 gastrointestinal stromal tumors for a proposed clinical staging system. *Arch Surg.* 2009;144:670–8.

23. Gold JS, Gönen M, Gutiérrez A, et al. Development and validation of a prognostic nomogram for recurrence-free survival after complete surgical resection of localised primary gastrointestinal stromal tumour: a retrospective analysis. *Lancet Oncol.* 2009;10:1045–52.

24. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* 1982;247:2543–6.

25. Dematteo RP, Ballman KV, Antonescu CR, et al; American College of Surgeons Oncology Group (ACOSOG) Intergroup Adjuvant GIST Study Team. Adjuvant imatinib mesylate after resection of localised, primary gastrointestinal stromal tumour: a randomised,double-blind, placebo-controlled trial. *Lancet.* 2009; 373(9669):1097–104.

26. Tanimine N, Tanabe K, Suzuki T, Tokumoto N, Ohdan H. Prognostic criteria in patients with gastrointestinal stromal tumors: a single center experience retrospective analysis. *World J Surg Oncol.* 2012;10:43.

27. Cohen MH, Cortazar P, Justice R, Pazdur R. Approval summary: imatinib mesylate in the adjuvant treatment of malignant gastrointestinal stromal tumors. *Oncologist.* 2010;15:300–7.

28. Joensuu H, Eriksson M, Sundby Hall K, et al. One vs three years of adjuvant imatinib for operable gastrointestinal stromal tumor: a randomized trial. *JAMA.* 2012;307:1265–72.

29. Patel S. Navigating risk stratification systems for the management of patients with GIST. *Ann Surg Oncol.* 2011;18:1698–704.

30. Joensuu H. Predicting recurrence-free survival after surgery for GIST. *Lancet Oncol.* 2009;10:1025.