Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts

Heikki Joensuu, Aki Vehtari, Jaakko Riihimäki, Toshirou Nishida, Sonja E Steigen, Peter Brabec, Lukas Plank, Bengt Nilsson, Claudia Cirilli, Chiara Braconi, Andrea Bordoni, Magnus K Magnusson, Zdenek Linke, Jozef Sufliarsky, Massimo Federico, Jon G Jonasson, Angelo Paolo Dei Tos, Piotr Rutkowski

Summary

Background The risk of recurrence of gastrointestinal stromal tumour (GIST) after surgery needs to be estimated when considering adjuvant systemic therapy. We assessed prognostic factors of patients with operable GIST, to compare widely used risk-stratification schemes and to develop a new method for risk estimation.

Methods Population-based cohorts of patients diagnosed with operable GIST, who were not given adjuvant therapy, were identified from the literature. Data from ten series and 2560 patients were pooled. Risk of tumour recurrence was stratified using the National Institute of Health (NIH) consensus criteria, the modified consensus criteria, and the Armed Forces Institute of Pathology (AFIP) criteria. Prognostic factors were examined using proportional hazards and non-linear models. The results were validated in an independent centre-based cohort consisting of 920 patients with GIST.

Findings Estimated 15-year recurrence-free survival (RFS) after surgery was 59·9% (95% CI 56·2–63·6); few recurrences occurred after the first 10 years of follow-up. Large tumour size, high mitosis count, non-gastric location, presence of rupture, and male sex were independent adverse prognostic factors. In receiver operating characteristics curve analysis of 10-year RFS, the NIH consensus criteria, modified consensus criteria, and AFIP criteria resulted in an area under the curve (AUC) of 0·79 (95% CI 0·76–0·81), 0·78 (0·75–0·80), and 0·82 (0·80–0·85), respectively. The modified consensus criteria identified a single high-risk group. Since tumour size and mitosis count had a non-linear association with the risk of GIST recurrence, novel prognostic contour maps were generated using non-linear modelling of tumour size and mitosis count, and taking into account tumour site and rupture. The non-linear model accurately predicted the risk of recurrence (AUC 0·88, 0·86–0·90).

Interpretation The risk-stratification schemes assessed identify patients who are likely to be cured by surgery alone. Although the modified NIH classification is the best criteria to identify a single high-risk group for consideration of adjuvant therapy, the prognostic contour maps resulting from non-linear modelling are appropriate for estimation of individualised outcomes.

Funding Academy of Finland, Cancer Society of Finland, Sigrid Juselius Foundation and Helsinki University Research Funds.

Introduction

Gastrointestinal stromal tumours (GIST) have varying malignancy potential. Micro-GISTs, with a diameter of less than 1 cm, occur in roughly 30% of the middle-aged and elderly general population.[1,2] Micro-GISTs have almost no malignancy potential, even though many of these tumours harbour an activating mutation in *KIT* or *PDGFRA*, which are considered the key drivers of GIST molecular pathogenesis.[3] Large GISTs and GISTs with a high mitosis count have a high recurrence rate, with metastases typically in the liver and abdominal cavity. Metastatic GISTs are often lethal despite treatment with tyrosine kinase inhibitors.[4]

Estimation of the risk of recurrence is important in the management of operable GIST. Adjuvant imatinib increases the time to GIST recurrence,[5,6] and patients with a high risk of recurrence have longer survival with 3 years of adjuvant imatinib therapy than with 1 year.[6] Although adjuvant imatinib is generally well tolerated, nearly all patients report some adverse effects.[6] Since many GIST patients are likely cured by surgery and might not benefit further from adjuvant treatment, risk stratification is one of the challenges in the management of operable GIST.

A few risk-stratification schemes are available for operable GIST.[7–11] Tumour size, mitosis count, and tumour site are considered established risk factors for recurrence. Patients with gastric GIST generally have more favourable prognosis than those with intestinal GIST.[7] Tumour rupture either spontaneously or at surgery is associated with a high risk of recurrence,[12–14] but whether rupture is an independent risk factor is controversial. Some mutations, such as deletions involving *KIT* exon 11 at codons 557–558, and many tumour biological factors, are associated with unfavourable outcome, but they are not considered established independent risk factors and are not included in the current risk-stratification schemes.[15]

Lancet Oncol 2012; 13: 265–74

Published Online December 7, 2011 DOI:10.1016/S1470-2045(11)70299-6

See Comment page 220

Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland (Prof H Joensuu MD); **Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland** (A Vehtari DSci, J Riihimäki MSc); **Osaka Police Hospital, Osaka, Japan** (T Nishida MD); **University Hospital of Northern Norway and University of Tromsø, Tromsø, Norway** (S E Steigen MD); **Institute of Biostatistics and Analyses, Masaryk University, Czech Republic** (P Brabec MSci); **Department of Pathology, Jessenius Medical Faculty of Comenius University and University Hospital, Martin, Slovak Republic** (Prof L Plank MD); **Department of Surgery, Sahlgrenska University Hospital, Gothenburg, Sweden** (B Nilsson MD); **Registro Tumori della provincia di Modena, Modena, Italy** (C Cirilli BSc); **Centro Regionale di Genetica Oncologica, Oncologia Medica, Ancona, Italy** (C Braconi MD); **Ticino Cancer Registry, Insitute of Pathology South of Switzerland, Locarno, Switzerland** (A Bordoni MD); **Department of Laboratory Hematology and Pathology, Landspitali University Hospital, Reykjavik, Iceland** (M K Magnusson MD); **Department of Internal Medicine, National Cancer Institute, Bratislava, Slovak Republic** (J Sufliarsky MD); **Faculty Hospital Motol, Radiotherapeutical-Oncological Department, Prague, Czech Republic** (Z Linke MD); **Department of Oncology and Hematology, University of Modena and**

Reggio Emilia, Modena, Italy (M Federico MD)**; Department of Pathology, Landspitali University Hospital and the Faculty of Medicine, University of Iceland, Reykjavik, Iceland** (J G Jonasson MD)**; General Hospital of Treviso, Treviso, Italy** (A P Dei Tos, MD)**; and Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland** (P Rutkowski MD)

Correspondence to:
Prof Heikki Joensuu, Department of Oncology, Helsinki University Central Hospital, Haartmaninkatu 4, POB 180, FIN-00029 Helsinki, Finland
heikki.joensuu@hus.fi

Although GIST is a common type of sarcoma with an annual incidence of 10–15 cases per million,[16–19] it is a recently defined tumour entity, so research on risk stratification has suffered from a lack of large, unselected series with long follow-up. The relative accuracy of widely used risk stratification schemes[7,9,11] in predicting outcome is unknown. These schemes are based on categories of tumour size and mitosis count, but it is unknown whether the chosen cutoff values are optimum. We hypothesised that the performance of established risk factors and the most widely used risk-stratification schemes can be best studied in population-based cohorts of patients with GIST, to avoid selection bias. To create a large database, we pooled individual patient data on the key prognostic factors from population-based series identified from the literature. We also developed a novel risk-stratification method where tumour size and mitosis count are treated as continuous non-linear variables.

Methods

Patients

This was an observational cohort study based on published population-based series of patients with operable GIST who did not receive adjuvant systemic therapy after surgery.[12,13,16–22] The study objectives were to assess key prognostic factors for recurrence-free survival (RFS) in operable GIST, to compare commonly used risk-stratification schemes, and to develop a new method for risk stratification. The study protocol was approved by an institutional review committee of the Helsinki University Central Hospital.

See Online for appendix

For the **study protocol**, see http://www.hus.fi/huchtrialprotocols

Eligible patients were required to have tumour morphology compatible with GIST and positive immunostaining for the KIT protein. Roughly 5% of GISTs do not stain for the KIT protein in immunohistochemistry,[3] and in such cases the diagnosis was considered confirmed if a mutation was found in either the *KIT* or *PDGFRA* genes. Only tumours that were removed macroscopically completely at surgery were eligible. We excluded individuals for whom date of diagnosis, age, or sex was unknown, and patients who had more than one GIST, recurrent GIST, or detectable metastases at the time of the diagnosis. Staging examinations were done according to the local practice.[12,13,16–22] Patients who had received adjuvant or neoadjuvant therapy were excluded.

Procedures

Population-based series of GIST published between Jan 1, 2000, and Jan 1, 2010, were identified by a search of PubMed and Scopus (including Embase) using the search terms "gastrointestinal stromal tumour", "GIST", and "population-based". We also reviewed the reference lists of relevant publications. From 148 series identified, we excluded those that were judged to be centre-based, those with fewer than 50 patients, series that focused on a GIST subgroup, and trials and studies that did not include relevant data on prognostic factors (see appendix). We contacted one or more of the investigators from 12 potentially eligible series. Two series were then excluded because the study database was no longer available,[23] or the study was published only in abstract form before Jan 1, 2010, and a substantial

	Time period	Number of patients	Sex: male	Median age at diagnosis in years (range)	Patients or tumours with data available					
					Tumour size	Mitotic count	Tumour site	Tumour rupture	RFS	Overall survival
Population-based series										
Modena, Italy[17]	1988–2010	157	90 (57·3%)	67 (25–90)	148 (94·3%)	148 (94·3%)	157 (100%)	0	156 (99·4%)	157 (100%)
Iceland[18]	1990–2003	50	29 (58·0%)	68 (24–89)	49 (98·0%)	50 (100%)	50 (100%)	49 (98·0%)	50 (100%)	50 (100%)
South Switzerland[20]	1999–2009	63	36 (57·1%)	67 (31–96)	61 (96·8%)	57 (90·5%)	63 (100%)	0	62 (98·4%)	63 (100%)
Ancona, Italy[21]	1987–2006	72	35 (48·6%)	62 (30–92)	72 (100%)	72 (100%)	72 (100%)	0	72 (100%)	72 (100%)
Western Sweden[16]	1971–2001	231	116 (50·2%)	68 (19–92)	231 (100%)	231 (100%)	231 (100%)	0	231 (100%)	231 (100%)
Northern Norway[19]	1971–2003	457	238 (52·1%)	67 (23–94)	336 (73·5%)	457 (100%)	430 (94·1%)	0	0	457 (100%)
Poland[12]	1981–2010	580	271 (46·7%)	60 (9–89)	566 (97·6%)	528 (91·0%)	580 (100%)	543 (93·6%)	580 (100%)	580 (100%)
Osaka, Japan[13]	1972–2009	474	251 (53·0%)	63 (10–93)	465 (98·1%)	408 (86·1%)	474 (100%)	470 (99·2%)	474 (100%)	474 (100%)
Slovak Republic[22]	1996–2010	224	104 (46·4%)	62 (20–94)	211 (94·2%)	81 (36·2%)	224 (100%)	136 (60·7%)	0	149 (66·5%)
Czech Republic[22]	1993–2010	252	133 (52·8%)	60 (14–90)	239 (94·8%)	227* (90·1%)	251 (99·6%)	0	0	226 (89·7%)
Total	1971–2010	2560	1303 (50·9%)	63 (9–96)	2378 (92·9%)	2259† (88·2%)	2532 (98·9%)	1198 (46·8%)	1625 (63·5%)	2459 (96·1%)
Validation series										
Italy	1980–2000	920	520 (56·5%)	66 (12–95)	903 (98·2%)	920 (100%)	920 (100%)	0	920 (100%)	909 (98·8%)

Data are number (%) unless otherwise indicated. RFS=recurrence-free survival. *Only categorised data available. †Mitosis count available both as a continuous and categorical variable for 1773 tumours, and as a categorical variable only for 486 tumours in the pooled series.

Table 1: **Characteristics of the series**

	Pooled series (n=2560)	Validation series (n=920)	p value
Age (years)			
Median (range)	63 (9–96)	66 (12–95)	<0·0001*
<20	11 (0·4%)	2 (0·2%)	
<50	419 (16·4%)	113 (12·3%)	
Sex			
Male	1303 (50·9%)	520 (56·5%)	0·003†
Female	1257 (49·1%)	400 (43·5%)	
Tumour site			
Stomach	1429 (56·4%)	555 (60·3%)	0·228†
Small intestine	828 (32·7%)	273 (29·7%)	
Colon or rectum	149 (5·9%)	52 (5·7%)	
Other	126 (5·0%)	40 (4·3%)	
NA	28	0	
Tumour size (cm)			
Median (range)	5·5 (0·1–45·0)	5·5 (0·1–50·0)	0·143*
≤1·0	129 (5·4%)	71 (7·9%)	
1·1–2·0	163 (6·9%)	84 (9·3%)	
2·1–5·0	847 (35·6%)	279 (30·9%)	
5·1–10·0	749 (31·5%)	290 (32·1%)	
10·1–15·0	300 (12·6%)	117 (13·0%)	
>15·0	190 (8·0%)	62 (6·9%)	0·002†
NA	182	17	
Mitoses per 50 HPFs			
Median (range)	3 (0–276)	3 (0–232)	0·628*
<2	780 (36·7%)	320 (34·8%)	
2–5	628 (29·6%)	298 (32·4%)	
6–10	253 (11·9%)	103 (11·2%)	
>10	462 (21·8%)	199 (21·6%)	0·451†
NA	437‡	0	
Tumour rupture			
No	1127 (94·1%)	NA	
Yes	71 (5·9%)	NA	
NA	1362	..	
Mutated exon			
KIT exon 11	701 (65·4%)	NA	
KIT exon 9	66 (6·2%)	NA	
KIT exon 13	10 (0·9%)	NA	
KIT exon 17	7 (0·7%)	NA	
PDGFRA exon 12	15 (1·4%)	NA	
PDGFRA exon 18	70 (6·5%)	NA	
Wild type	198 (18·5%)	NA	
Other§	5 (0·5%)	NA	
NA	1488	..	

Data are number (%), unless otherwise indicated. HPF=high power field of the microscope. NA=not available. *p value for difference between medians; Mann-Whitney test. †p value for difference between categories; χ^2 test. ‡Includes 136 cases where mitosis count was available, categorised as <5 mitoses per 50 HPFs (from the Czech Republic series). §*KIT* and *PDGFRA* exons.

Table 2: **Characteristics of patients and tumours**

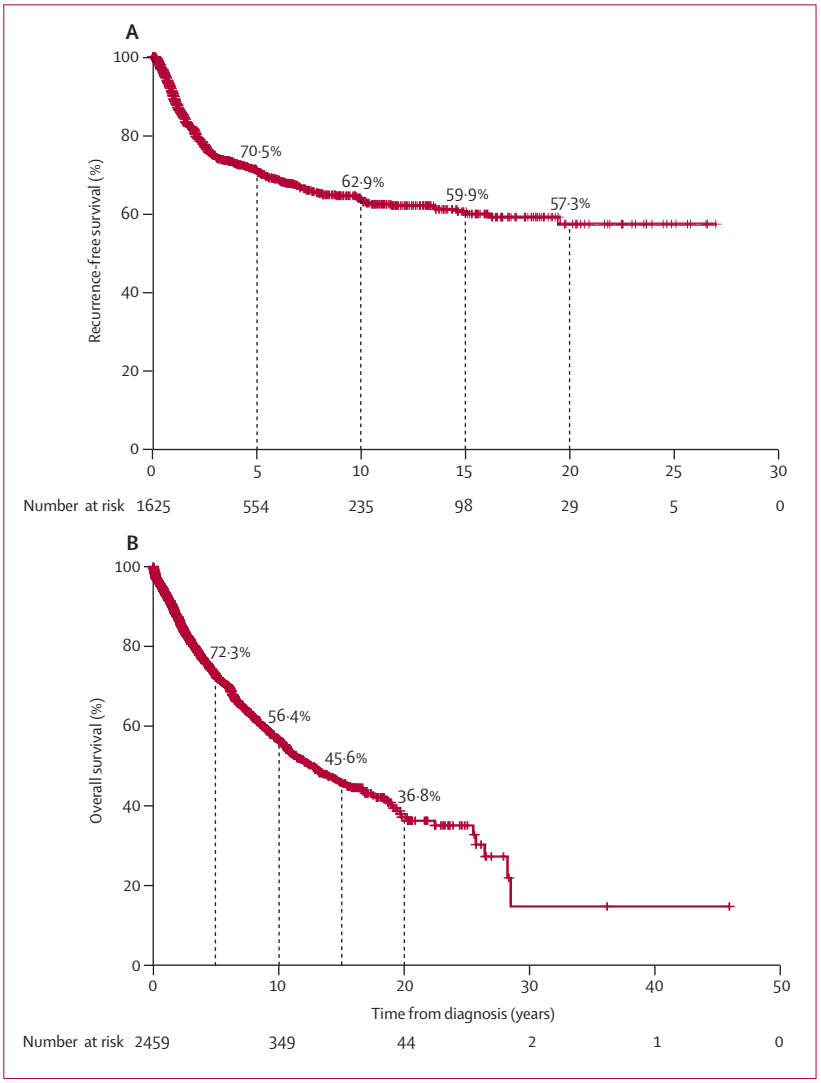


Figure 1: **Recurrence-free survival (A) and overall survival (B) in the pooled series**

proportion of the patients received adjuvant imatinib.[24] The remaining ten series form the basis of the present study (table 1).[12,13,16–22]

Each patient was assigned a study code to protect their identity before data transfer. From the 2688 patients with data transferred, we excluded 128 patients who did not fulfil eligibility criteria, leaving 2560 patients in the study. An investigator from each study approved the final dataset before it was merged into the pooled dataset. Data from an independent validation series consisting of 920 patients who had KIT-positive operable GIST, diagnosed at 35 hospitals in Italy from 1980–2000, was obtained from one of the study investigators.[25] Criteria for inclusion and exclusion of patients in the validation series were similar to the criteria used for the pooled population-based series.

Tumours were classified using the National Institute of Health (NIH) consensus criteria,[9] the modified NIH consensus criteria,[11,15] and the Armed Forces Institute of Pathology (AFIP) criteria[7] (appendix). In the NIH consensus classification, tumours are stratified into four

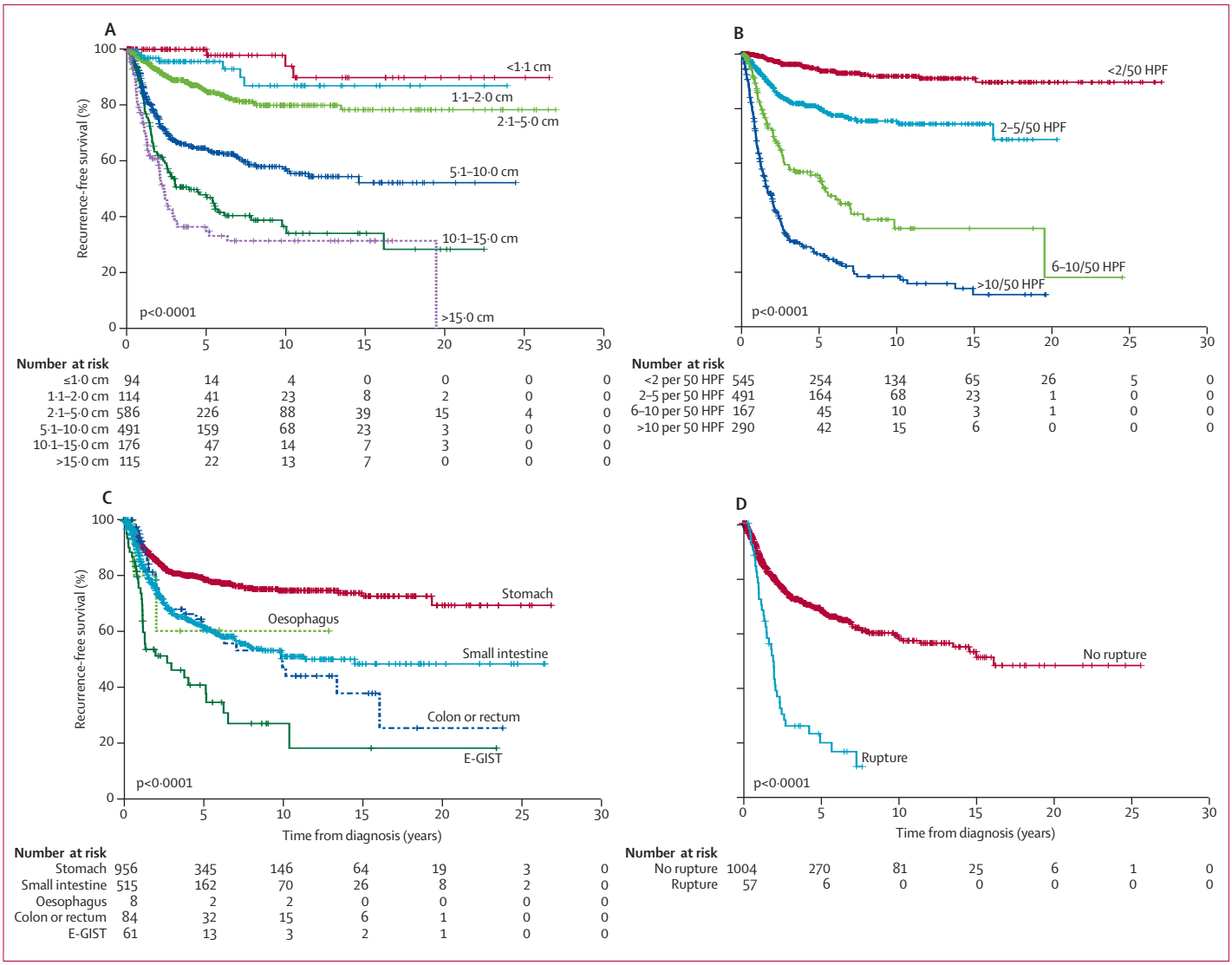
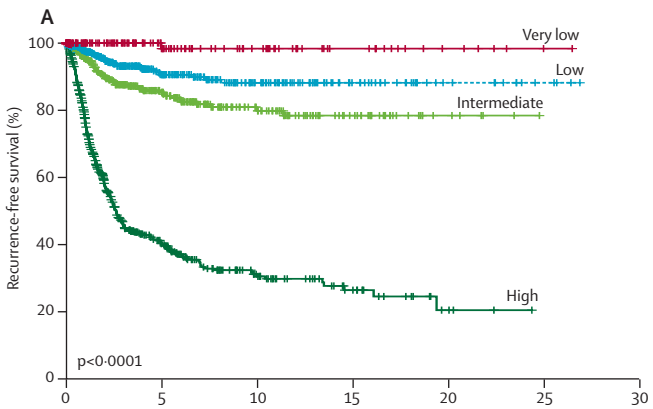
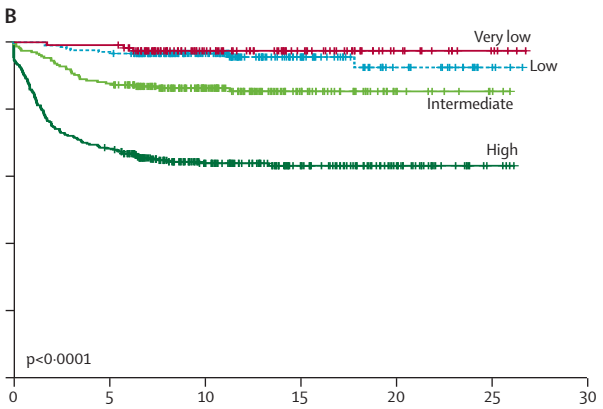


Figure 2: **Recurrence-free survival by tumour size (A), mitosis count (B), site (C), and rupture (D) from univariable analysis of the pooled dataset**
HPF=high power field of the microscope. E-GIST=extragastrointestinal stromal tumour (GIST that arises outside the gastrointestinal tract).

risk groups based on two factors—size and mitosis count (very low risk: size <2·0 cm and mitosis count <5 per 50 high power fields [HPFs] of the microscope; low risk: size 2·0–4·9 cm and mitosis count <5 per 50 HPFs; intermediate risk: size ≤5·0 cm and mitosis count 6–10 per 50 HPFs, or size 5·0–10·0 cm and mitosis count <5 per 50 HPFs; high risk: size >5·0 cm and mitosis count >5 per 50 HPFs, or size >10·0 cm with any mitosis count, or mitosis count >10 per 50 HPFs irrespective of size).[9] Since the NIH consensus criteria do not define how to classify tumours with exactly five mitoses per 50 HPFs, we included these tumours in the very low-risk, low-risk, or intermediate risk categories depending on size. The modified NIH classification differs from the NIH consensus classification by considering ruptured GIST as a high-risk tumour irrespective of other features, and by including in the high-risk category small (≤5·0 cm) non-gastric tumours with frequent mitoses (>5 per 50 HPFs), and intermediate size (5·1–10·0 cm) non-gastric tumours with few mitoses (≤5 per 50 HPFs).[11,15] The AFIP classification stratifies tumours with few mitoses (≤5 per 50 HPFs) into four categories based on size (group 1: <2·0 cm; group 2: 2·1–5·0 cm; group 3a: 5·1–10·0 cm; group 3b: >10·0 cm) and tumours with frequent mitoses (>5 per 50 HPFs) into four groups (group 4: <2·0 cm; group 5: 2·1–5·0 cm; group 6a: 5·1–10·0 cm; group 6b: >10·0 cm). The AFIP

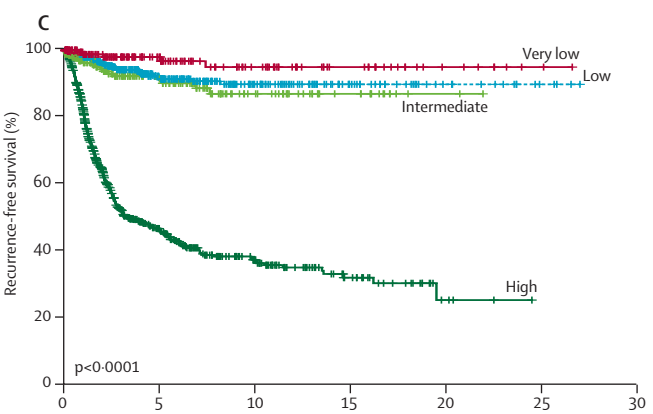
Figure 3: **Recurrence-free survival stratified by the National Institute of Health (NIH) consensus criteria (A, B), the modified NIH consensus criteria (C, D), and the Armed Forces Institute of Pathology (AFIP) criteria (E, F)**
Panels A, C, E show the pooled population-based series; panels B, D, F show the validation series.

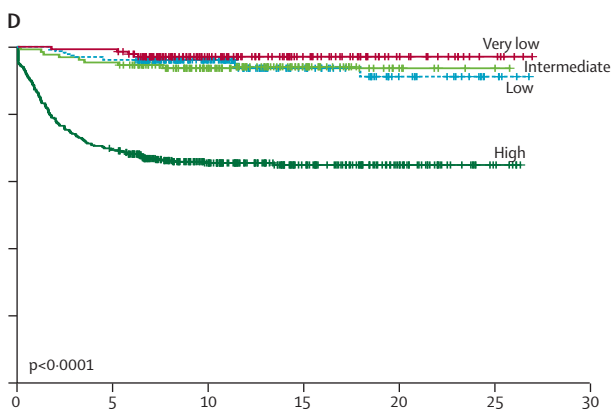




Number at risk							
Very low risk	130	59	33	16	6	1	0
Low risk	429	164	74	32	12	4	0
Intermediate risk	395	161	71	29	6	0	0
High risk	554	122	47	20	4	0	0

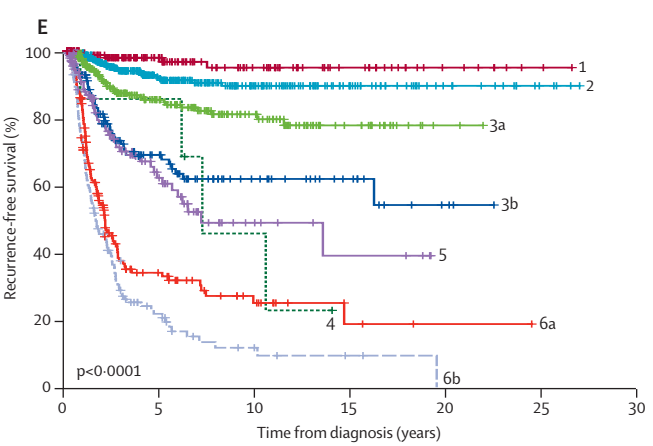
120	119	68	39	15	6	0
210	204	116	52	20	4	0
237	208	114	45	12	3	0
342	233	131	71	32	4	0

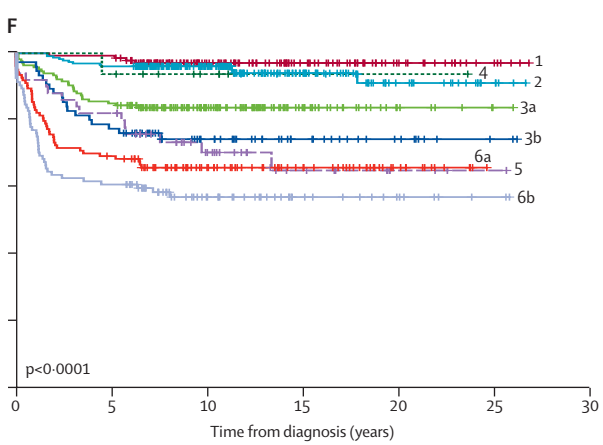




Number at risk							
Very low risk	180	77	43	22	8	1	0
Low risk	435	174	79	34	14	4	0
Intermediate risk	205	86	34	14	2	0	0
High risk	694	171	69	27	4	0	0

137	136	79	44	18	6	0
229	220	122	52	19	5	0
128	122	68	27	7	1	0
415	286	160	84	35	5	0





Number at risk							
Group 1	186	78	44	22	8	1	0
Group 2	430	173	78	34	14	5	0
Group 3a	278	114	54	20	2	0	0
Group 3b	130	49	23	12	3	0	0
Group 4	8	5	2	0	0	0	0
Group 5	113	35	8	4	0	0	0
Group 6a	177	30	11	3	1	0	0
Group 6b	145	17	4	2	1	0	0

139	138	80	45	19	6	0
229	220	122	52	19	5	0
166	141	75	30	7	1	0
75	58	34	18	9	2	0
16	15	10	5	1	0	0
50	41	22	10	3	1	0
123	84	52	31	12	0	0
104	63	31	14	7	2	0

classification also takes into account tumour site (gastric, duodenal, jejunal and ileal, or rectal), and the risks involved with each group and site have been summarised.[7]

Statistical analysis

RFS (the primary endpoint) was calculated from the date of surgery or date of histological diagnosis to the date of first tumour recurrence, censoring patients alive at the time of data collection and those who died without GIST recurrence on the date of death. Overall survival was calculated from the date of surgery or the date of diagnosis to the date of death, censoring patients who were alive. Survival between groups was compared using the Kaplan-Meier life-table method and a non-stratified Cox proportional hazards model or log-rank test. Prognostic factors were compared using a Cox proportional hazards model. Frequency tables were analysed using the χ^2 test. Non-normal distributions were compared with the Mann-Whitney *U* test. p values were two-sided and were not adjusted for multiple testing. Analyses were done with SPSS version 17.0.

Non-linear modelling was done using a modified Cox proportional hazard model, where the log-linear predictor was replaced by a logarithmic Gaussian process (GP), and the piecewise log-constant baseline hazard was smoothed using another GP, producing a non-parametric GP-Cox model.[26,27] This model is suitable for assessing non-linear effects and implicit interactions between covariates. The conventional Cox model is log-linear, and so non-linear effects can only be examined after covariate categorisation; however, the GP-Cox model is suitable for modelling of smooth non-linear effects and does not require covariate categorisation (appendix). The Bayesian inference was used, and the model was implemented using GPstuff software.[28]

For **GPstuff software**, see http://www.lce.hut.fi/research/mm/gpstuff/

Receiver operating characteristics (ROC) curves and the corresponding areas under the curve (AUC) for the compared models were calculated using ten-fold cross-validation to simulate predictive accuracy in an unseen population.[29] An independent replicate dataset was used to verify generalisation of the model.

Role of the funding source

The funding sources had no role in the study design, data collection, analysis, or interpretation, or writing of the manuscript. HJ, AV, and JR had full access to the raw data of the study database. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

Information about primary tumour size, mitosis count, site, and rupture were available in the pooled dataset for 2378, 2259, 2532, and 1198 patients, respectively; data on RFS and overall survival were available for 1625 and 2459 patients, respectively (table 1). The median follow-up time for patients alive at the time of data collection was 4·0 years (range 0–45·8).

The median age at the time of GIST diagnosis was 63 years (range 9–96); only 11 (0·4%) of 2560 patients were younger than 20 years (table 2). There was a slight male predominance (50·9% [1303 of 2560]), and most tumours were located in the stomach (56·4% [1429 of 2532]; the smallest proportion of gastric tumours was 49·0% [123 of 251] in the Czech Republic series, and the highest proportion was 69·2% [328 of 474] in the Osaka series). Median tumour diameter at the time of the diagnosis was 5·5 cm (range 0·1–45·0; the smallest median size was 3·7 cm [0·1–20·0; Iceland] and the largest median size was 7·5 cm [0·1–23·0; Ancona, Italy]), and 490 (20·6%) of 2378 patients had tumours which were larger than 10 cm. The median number of mitoses per 50 HPFs was three (range 0–276; this median ranged from two to four in all series except for the Iceland series, where it was 0·5). 71 (5·9%) of 1198 patients had tumour rupture either spontaneously or at surgery. *KIT* or *PDGFRA* mutations were detected in 787 (73·4%) and 87 (8·1%), respectively, of the 1072 tumours with mutation data available.

Compared with patients in the pooled series, patients in the validation series were older (median age at diagnosis was 66 years [range 12–95]; p<0·0001) and a higher proportion were male (56·5% [520 of 920]; p=0·003), but there was no difference in median tumour size, tumour sites, or the number of mitoses between the series (table 2). Median follow-up time for patients in the validation series was 9·4 years (range 4·8–26·9).

The estimated median overall survival time in the pooled series was 12·4 years (95% CI 10·8–14·0). Most GIST recurrences took place within the first 5 years of follow-up, and few tumours recurred after the first 10 years of follow-up, up to 19·4 years after surgery (figure 1). Estimated 5-year, 10-year, and 15-year RFS were 70·5% (68·0–73·0), 62·9% (59·8–66·0), and 59·9% (56·2–63·6), respectively.

Tumour size and mitosis count were strongly associated with RFS in univariable analysis (figure 2). Compared with the smallest size category (diameter <1·1 cm), the categories 1·1–2·0 cm, 2·1–5·0 cm, 5·1–10·0 cm, 10·1–15·0 cm, and more than 15·0 cm were associated with a hazard ratio (HR) of 2·19 (95% CI 0·57–8·49; p=0·255), 4·45 (1·40–14·11; p=0·011), 12·84 (4·10–40·24; p<0·0001), 21·56 (6·82–68·12, p<0·0001), and 27·98 (8·79–89·06; p<0·0001), respectively. Compared with a very low tumour mitosis count (<2 per 50 HPFs), the HRs associated with a high count (>10 per 50 HPFs), moderate count (6–10 per 50 HPFs), and low count (2–5 per 50 HPFs) were 22·09 (14·98–32·58; p<0·0001), 11·10 (7·24–17·01; p<0·0001), and 3·78 (2·49–5·75; p<0·0001), respectively. Compared with intestinal GISTs, gastric GISTs were associated with better RFS (0·49, 0·35–0·67; p<0·0001), oesophageal and colorectal GISTs with similar RFS

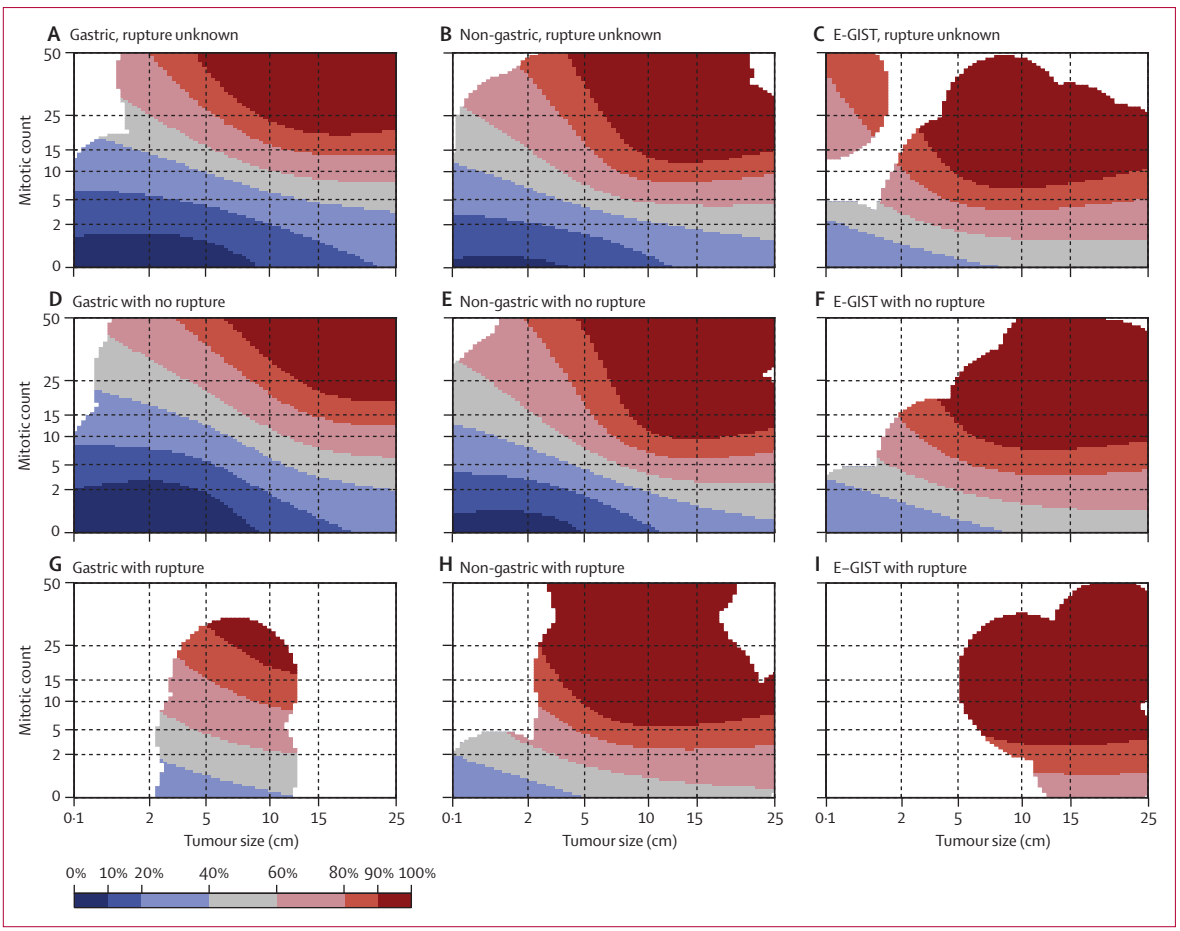


Figure 4: **Contour maps for estimating the risk of GIST recurrence after surgery**
The upper row maps are used when tumour rupture status is unknown (A,B,C), the middle row maps when the tumour has not ruptured (D,E,F), and the bottom row maps when tumour rupture has occurred (G,H,I). Red areas depict high risk, blue areas low risk, and white areas indicate lack of data. The percentages associated with each colour (key) indicate the probability of GIST recurrence within the first 10 years of follow-up after surgery. For example, the middle map of the far left column (D) shows that the 10-year risk of GIST recurrence of a patient diagnosed with a 10 cm gastric GIST with five mitoses per 50 high power fields (HPFs) of the microscope and no rupture is 20–40%. The 10-year risk associated with a similar tumour when the mitosis count is ten per 50 HPFs increases to 40–60%.
E-GIST=extragastrointestinal stromal tumour (arising outside the gastrointestinal tract).

(0·94, 0·31–2·85; p=0·908; and 1·05, 0·71–1·55; p=0·826, respectively) and GISTs located outside of the gastrointestinal tract (E-GISTs) with worse outcome (2·13, 1·44–3·15, p<0·0001). Women had slightly better RFS than men (0·90, 0·82–0·99; p=0·025).

When tumour size, site, mitosis count, rupture (each categorised as in figure 2), age at diagnosis (≤median *vs* >median), and sex were entered as covariates in a Cox model, tumour rupture (HR 1·65, 95% CI 1·10–2·46; p=0·015) and sex (1·38, 1·07–1·77; p=0·013) were independently associated with RFS, in addition to tumour size, site, and mitosis count (p<0·001 for each). When tumour rupture was excluded from the analysis (since data on rupture were often missing), results remained similar for the rest of the factors (data not shown).

The NIH consensus criteria, modified NIH criteria, and AFIP criteria were all strongly associated with RFS in the pooled dataset and the validation series (figure 3). In both datasets, the modified NIH criteria was the best criteria for identifying a single high-risk group, whereas the AFIP criteria produced subgroups with RFS varying from very good to unfavourable. The AFIP criteria's group 4 was small in both datasets (eight [0·5%] of 1467 patients in the pooled series and 16 [1·8%] of 902 in the validation series).

Increasing tumour size and mitosis count were associated with worsening RFS (figure 2). The effect of such factors on survival is traditionally estimated using log-linear modelling in a proportional hazard model, but both tumour size and mitosis count had a non-linear association with the logarithm of the hazard of GIST recurrence. This value increased more per one unit of measurement of the variable (per 1 cm or one mitosis per 50 HPFs) at small values (ie, when tumour size was small or mitosis count low) compared with large values

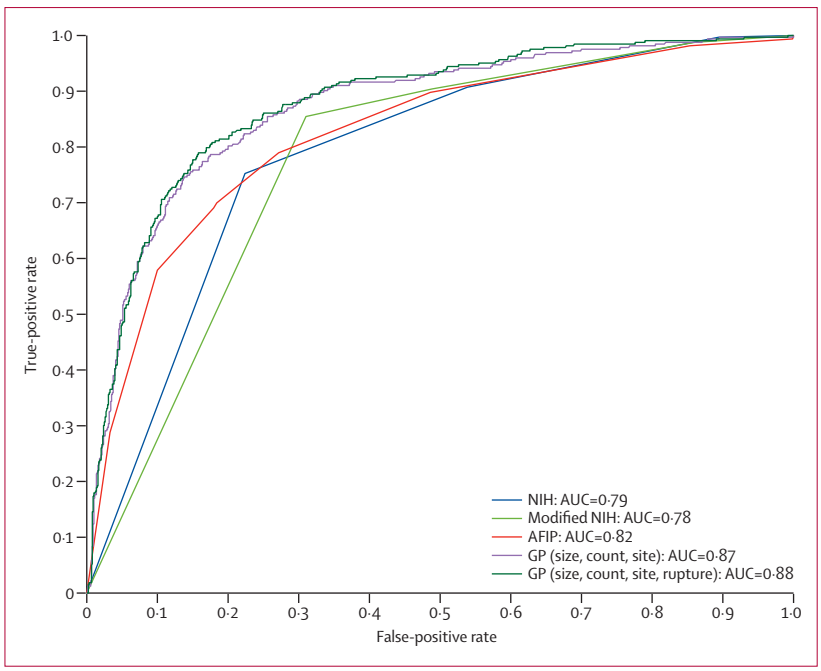


Figure 5: **Receiver operating characteristic (ROC) analysis of the risk of GIST recurrence during the first 10 years of follow-up after surgery**

GIST=gastrointestinal stromal tumour. AUC=area under the curve. NIH=National Institutes of Health. AFIP= Armed Forces Institute of Pathology. GP=Gaussian process (non-linear modelling of the data).

(appendix). When non-linear models (the GP-Cox models) were applied to the pooled data, prognostic heat maps (appendix) and contour maps (figure 4) were produced using tumour size, site, mitosis count, and rupture, and considering tumour size and mitosis count as continuous non-linear variables. Sex and age at diagnosis did not affect the predictive performance of the GP-Cox models.

The prognostic accuracy of the non-linear models and the conventional risk-stratification schemes was compared using ROC analyses. The non-linear models produced the best estimations for the risk of GIST recurrence. In estimating the 10-year risk of GIST recurrence, the AUC was larger for the non-linear model that included tumour rupture data (0·88, 95% CI 0·86–0·90) than for the NIH consensus classification criteria (0·79, 0·76–0·81; p<0·0001), modified consensus classification criteria (0·78, 0·75–0·80; p<0·0001), AFIP criteria (0·82; 0·80–0·85; p<0·0001) or the non-linear model that did not include tumour rupture data (0·87, 0·85–0·89; p=0·005; figure 5). In the validation series, where rupture data were unavailable, the non-linear model (AUC 0·80, 0·76–0·83) was slightly more accurate than the consensus criteria (0·76, 0·73–0·79; p=0·0003), modified consensus criteria (0·76, 0·73–0·79; p=0·001), or AFIP criteria (0·77, 0·73–0·80, p=0·003; appendix). The results remained similar when GIST recurrence based on either 5 or 15 years of follow-up was used as the endpoint (data not shown).

Discussion

We pooled data from GIST series that were considered by the investigators to be population-based, to estimate the fraction of patients cured by surgery, to assess the performance of standard risk-stratification schemes, and to develop a new method for estimation of the risk of GIST recurrence (panel). The results for this large cohort of patients with operable GIST suggest that most patients are cured by surgery and thus do not benefit from systemic adjuvant therapy. To our knowledge, the present study is the first to compare the NIH consensus criteria, modified consensus criteria, and AFIP criteria for risk stratification. The results are reassuring, since all schemes accurately predicted outcome. The AFIP criteria identified patients with widely different outcomes, whereas the modified consensus criteria identified a single subgroup with unfavourable prognosis, which might be helpful when considering the need for adjuvant therapy. The novel prognostic heat maps and contour maps were more accurate than these risk-stratification schemes in estimating the risk of GIST recurrence after surgery for patients who did not have more than one GIST, recurrent GIST, or detectable metastases at the time of the diagnosis.

Each population-based series contained data from a vast majority of all operable GISTs diagnosed within a defined geographical region and time period, and was pooled to obtain an unselected patient cohort that reflects GIST globally. Carrying out a similar study in the future might be increasingly challenging, since adjuvant imatinib modifies patient outcomes[5,6] and will probably be given often, making unselected cohorts of patients who receive surgery alone rare. To maintain an unselected cohort, we did not exclude a small minority of GISTs with a diameter of 1 cm or less, although some of these tumours might have been micro-GISTs with low malignancy potential.

Most recurrences of operable GIST take place within the first 5 years after surgery, but in accordance with prior data,[15] we found that GIST can occasionally recur later, sometimes 20 years after surgery. Prognosis of patients with GIST was better in the validation series than in the pooled population-based series. The reasons for this remain speculative, but case selection and completeness of reporting of GIST recurrences might have had a role. By contrast with the pooled series, all GISTs in the validation series were required to express the KIT protein; however, since KIT-negative GISTs are rare, this is unlikely to explain the difference.

Mitosis count is a strong prognostic factor in GIST, but it has limitations and its reliability is controversial. Identification of mitoses can be subjective, and the number detected depends on the tissue fixation time and the size of the field of view of the microscope.[30] Data on the latter were not typically available.[12,13,17,18,20–22] Although mitosis count was done by many pathologists and the methods used might have varied, the counts were

strongly associated with RFS and there was only moderate variation in the median numbers of counts between series. Mitosis count might be best provided as per 1 mm² of tumour tissue, rather than per 50 HPFs; it remains to be investigated how this would affect the estimated risk of GIST recurrence and risk stratification.

Little information is available regarding outcome of colorectal GISTs because of their rarity. A prognostic nomogram[10] considers these tumours to have a somewhat favourable outcome, slightly worse than that of gastric GISTs, whereas the present data suggest that their prognosis is not markedly different from that of small intestinal GISTs at 10 years. E-GISTs that arise in the abdominal cavity outside of the gastrointestinal tract were associated with the most unfavourable outcome, but some E-GISTs might have been metastases from an undetected primary tumour. The small number of oesophageal and colorectal GISTs prohibits making firm conclusions about their outcome.

Patients with ruptured GIST have poor outcome. Since large tumours with unfavourable histopathological features rupture most often, it is controversial whether tumour rupture is an independent prognostic factor.[11–15] Tumour rupture had an independent adverse effect on RFS, and its inclusion slightly improved the accuracy of the non-linear model. The numbers of tumours that ruptured spontaneously or as a result of manipulation at surgery were unavailable, which is a limitation of this study.

The type of *KIT* or *PDGFRA* mutation likely affects survival.[3,15] Thus far, mutation type has not been integrated into risk-stratification schemes, probably because a few hundred different single mutations have been identified in GIST and large enough series with adequate follow-up to address their effect on outcome are lacking. The *KIT* and *PDGFRA* mutation frequencies found in the pooled series are typical of GIST,[3] but mutation data were available only for a small proportion of tumours.

The available risk-stratification schemes categorise tumour size and mitosis count,[7–11] which probably results in a loss of prognostic information and causes abrupt changes in the estimated risks of recurrence when tumour size or mitosis count is close to a cutoff value. We noted a non-linear effect of tumour size and mitosis count on RFS, with a larger effect per one unit of measurement at low values compared with high values; to address this, we created prognostic heat maps and user-friendly contour maps for estimating risk of GIST recurrence. These maps might be useful in patient counselling.

We conclude that most patients with operable GIST are cured by surgery alone. Risk-stratification schemes fairly accurately estimate RFS, and their use might reduce the risk of overtreatment with systemic adjuvant therapy in patients with GIST that is likely cured by surgery. Low-risk patients have generally favourable outcomes and might not be candidates for adjuvant therapy. Tumour rupture is an independent prognostic factor in operable GIST. When the modified NIH classification criteria are used, patients with intermediate-risk GIST have as favourable RFS as those with low-risk GIST, leaving only a single high-risk group for consideration of adjuvant therapy. Models that address the continuous and non-linear nature of tumour size and mitosis count produced more accurate estimations for the risk of GIST recurrence than the schemes that categorise these variables, and the resulting prognostic maps are appropriate for estimation of individualised outcomes.

Panel: Research in context

Systematic review

We searched the PubMed and Scopus databases to identify published series on gastrointestinal stromal tumours (GIST) that included nearly all patients diagnosed with GIST within a defined geographical region and a time period. We used the search terms "gastrointestinal stromal tumour", "GIST", and "population-based". We limited our search to reports published between January 1, 2000, and January 1, 2010, since GIST was not a generally recognised tumour entity before 2000, and adjuvant imatinib is now being used increasingly, often modifying the clinical course. We contacted the investigators, reviewed the data, and pooled the individual patient data from ten eligible population-based series from nine countries. The resulting database with individual data from 2560 patients with operable GIST allowed us to examine prognostic factors, compare the performance of the most commonly used risk-stratification schemes, and to develop a new prognostic model for patients with operable GIST.

Interpretation

The present unselected, multinational data show that roughly half of patients with operable GIST are cured by surgery alone and thus do not benefit from adjuvant imatinib. The data confirm that these patients can be identified relatively well with one of the commonly used risk-stratification schemes. We found similar accuracy for the different risk-stratification schemes. These schemes categorise two key prognostic variables, tumour size and mitosis count, whereas the most accurate estimates for risk of GIST recurrence were achieved when tumour size and mitosis count were modelled as continuous non-linear variables, and presence of tumour rupture and tumour site were also taken into account. Novel prognostic heat maps and contour maps, produced by non-linear modelling of the data, are provided in the article and appendix; they are appropriate for estimation of individualised outcomes.

Contributors

HJ designed the study, wrote the study protocol, did the literature search, and compiled the study database. TN, SES, PB, LP, BN, CC, CB, AB, MKM, ZL, JS, FM, JGJ, APDT, and PR collected and provided the original data. TN, SES, PB, LP, BN, CC, CB, AB, MKM, FM, JGJ, and PR reviewed the transmitted data. HJ, AV, and JR did the statistical analyses. AV and JR designed the non-linear model. HJ drafted the manuscript, and AV and JR participated in writing. All authors reviewed and commented on the manuscript, and approved its final submission.

Conflicts of interest

APDT has received a research grant from Novartis. All other authors declared no conflicts of interest.

Acknowledgments

This study was supported by Academy of Finland (grants 218068, 218248, 129230, and 131449), Cancer Society of Finland, Sigrid Juselius Foundation, and Helsinki University Research Funds (grant TYH7206). Novartis Oncology provided financial support for the GIST registry in the Czech Republic and the Slovak Republic.

References

1 Kawanowa K, Sakuma Y, Sakurai S, et al. High incidence of microscopic gastrointestinal stromal tumors in the stomach. *Hum Pathol* 2006; **37:** 1527–35.

2 Agaimy A, Wünsch PH, Hofstaedter F, et al. Minute gastric sclerosing stromal tumors (GIST tumorlets) are common in adults and frequently show c-KIT mutations. *Am J Surg Pathol* 2007; **31:** 113–20.

3 Rubin BP, Heinrich MC, Corless CL. Gastrointestinal stromal tumour. *Lancet* 2007; **369:** 1731–41.

4 Blanke CD, Demetri GD, von Mehren M, et al. Long-term results from a randomized phase II trial of standard- versus higher-dose imatinib mesylate for patients with unresectable or metastatic gastrointestinal stromal tumors expressing KIT. *J Clin Oncol* 2008; **26:** 620–25.

5 DeMatteo RP, Ballman KV, Antonescu CR, et al. Adjuvant imatinib mesylate after resection of localised, primary gastrointestinal stromal tumour: a randomised, double-blind, placebo-controlled trial. *Lancet* 2009; **373:** 1097–104.

6 Joensuu H, Eriksson M, Hartmann J, et al. Twelve vs. 36 months of adjuvant imatinib as treatment of operable GIST with a high risk of recurrence: final results of a randomized trial (SSGXVIII/AIO). *Proc Am Soc Clin Oncol* 2011; **29:** abstr LBA1.

7 Miettinen M, Lasota J. Gastrointestinal stromal tumors: pathology and prognosis at different sites. *Semin Diagn Pathol* 2006; **23:** 70–83.

8 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:** 670–78.

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

10 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.

11 Rutkowski P, Bylina E, Wozniak A, et al. Validation of the Joensuu risk criteria for primary resectable gastrointestinal stromal tumour—the impact of tumour rupture on patient outcomes. *Eur J Surg Oncol* 2011; **37:** 890–96.

12 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.

13 Takahashi T, Nakajima K, Nishitani A, et al. An enhanced risk-group stratification system for more practical prognostication of clinically malignant gastrointestinal stromal tumors. *Int J Clin Oncol* 2007; **12:** 369–74.

14 Hohenberger P, Ronellenfitsch U, Oladeji O, et al. Pattern of recurrence in patients with ruptured primary gastrointestinal stromal tumour. *Br J Surg* 2010; **97:** 1854–59.

15 Joensuu H. Risk stratification of patients diagnosed with gastrointestinal stromal tumor. *Hum Pathol* 2008; **39:** 1411–19.

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:** 821–29.

17 Mucciarini C, Rossi G, Bertolini F, et al. Incidence and clinicopathologic features of gastrointestinal stromal tumors: a population-based study. *BMC Cancer* 2007; **7:** 230.

18 Tryggvason G, Gislason H, Magnusson MK, Jonasson JG. Gastrointestinal stromal tumors in Iceland, 1990–2003: the Icelandic GIST study, a population based incidence and pathologic risk stratification study. *Int J Cancr* 2005; **117:** 289–93.

19 Steigen SE, Eide TJ, Wasag B, Lasota J, Miettinen M. Mutations in gastrointestinal stromal tumors—a population-based study from Northern Norway. *APMIS* 2007; **115:** 289–98.

20 Mazzola P, Spitale A, Banfi S, Mazzucchelli L, Frattini M, Bordoni A. Epidemiology and molecular biology of gastrointestinal stromal tumors (GISTs): a population-based study in the South of Switzerland, 1999–2005. *Histol Histopathol* 2008; **23:** 1379–86.

21 Braconi C, Bracci R, Bearzi I, et al. KIT and PDGFRalpha mutations in 104 patients with gastrointestinal stromal tumors (GISTs): a population-based study. *Ann Oncol* 2008; **19:** 706–10.

22 Brabec P, Sufliarsky J, Linke Z, et al. A whole population study of gastrointestinal stromal tumors in the Czech Republic and Slovakia. *Neoplasma* 2009; **56:** 459–64.

23 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:** 445–49.

24 Emile Francois J, Brahimi S, Coindre J, et al. Molecular epidemiology of GISTs: incidence of PDGFRA and KIT exon 9 mutations in the large French population-based study molecGIST. *Proc Am Soc Clin Oncol* 2009; **27:** abstr 10535.

25 Rossi S, Miceli R, Messerini L, et al. Natural history of imatinib-naive GISTs: a retrospective analysis of 929 cases with long-term follow-up and development of a survival nomogram based on mitotic index and size as continuous variables. *Am J Surg Pathol* 2011; **11:** 1646–56.

26 Rasmussen C, Williams C. Gaussian processes for machine learning. Cambridge: MIT Press, 2006.

27 Ibrahim JG, Chen M-H, Sinha D. Bayesian survival analysis. New York: Springer, 2001.

28 Vanhatalo J, Riihimäki J, Hartikainen J, Vehtari A. Bayesian modeling with Gaussian processes using the MATLAB toolbox GPstuff. http://www.lce.hut.fi/research/mm/gpstuff/ (accessed Aug 12, 2011).

29 Vehtari A, Lampinen J. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput* 2002; **14:** 2439–68.

30 Gal R, Rath-Wolfson L, Rosenblatt Y, Halpern M, Schwartz A, Koren R. An improved technique for mitosis counting. *Int J Surg Pathol* 2005; **13:** 161–65.