

## Performance of a new clinical grading system for chronic graft-versus-host disease: a multicenter study

Görgün Akpek, Stephanie J. Lee, Mary E. Flowers, Steven Z. Pavletic, Mukta Arora, Shing Lee, Steven Piantadosi, Katherine A. Guthrie, James C. Lynch, Alessandra Takatu, Mary M. Horowitz, Joseph H. Antin, Daniel J. Weisdorf, Paul J. Martin, and Georgia B. Vogelsang

We recently reported 3 risk factors (RFs) at diagnosis of chronic graft-versus-host disease (cGVHD) that were significantly associated with increased nonrelapse mortality. These included extensive skin involvement (ESI), thrombocytopenia (TP), and progressive type of onset (PTO). The hazard ratio (HR) for mortality of the patients with prognostic score (PS) between 0 and 2 (intermediate-risk; 1 RF) compared to those with PS 0 (favorable-risk; 0 RF) was 3.7 (95% CI, 1.4, 9.3); the HR for patients with PS equal to or more than 2 (high-risk; > 1 RF) compared with intermediate-risk group was 6.9 (3.8, 12.4). A rare presentation of TP and PTO without ESI yielded a PS of 1.8 (intermediate-risk). This paper reports the performance of the

prognostic model and the individual RFs using data from an additional 1105 patients from University of Nebraska (n = 60), International Bone Marrow Transplantation Registry (n = 708), Fred Hutchinson Cancer Research Center (n = 188), and University of Minnesota (n = 149). The extent of skin involvement was quantified in 3 cohorts using the available data collected in different formats before the analysis. Although the HR for mortality of the patients in the intermediate-risk group versus those in the favorable-risk group ranged from 2.3 to 8.9 across the centers, it was between 1.6 to 6.9 for patients in the high-risk group versus those in the intermediate-risk group. Although TP itself was uni-

formly associated with increased risk of mortality across all test samples, ESI and PTO showed statistically significant associations with mortality in 1 and 2 cohorts, respectively. In conclusion, the model was predictive of cGVHD-specific survival, but the mortality hazard associated with ESI was lower in each of these test samples compared with the learning sample. Although the new clinical grading based on the model is promising because of its utility across multiple independent data sets, prospective validation is needed. (*Blood*. 2003;102:802-809)

© 2003 by The American Society of Hematology

### Introduction

Chronic graft-versus-host disease (cGVHD) remains the most common late complication of allogeneic stem cell transplantation (SCT) with reported incidences ranging from 30% to 80%.<sup>1</sup> Despite improvements in the prevention of acute GVHD, only T-cell depletion<sup>2</sup> has resulted in a concomitant decrease in the incidence of cGVHD. The high incidence of cGVHD partly relates to the increase in the fraction of patients surviving transplantation complications as well as the recent changes in clinical SCT methods. SCT using donors other than HLA-identical siblings and, recently, the use of blood as a source of hematopoietic stem cells are increasing.<sup>3</sup> Donor lymphocyte infusion whether given after nonmyeloablative SCT or for relapse is also reported to be associated with severe acute and chronic GVHD and sometimes with unusual GVHD presentations.<sup>4</sup> Although cGVHD may have beneficial graft-versus-leukemia/lymphoma effects, large, observational studies identify cGVHD as the leading cause of nonrelapse deaths occurring more than 2 years after transplantation.<sup>5</sup> Mortality related to cGVHD is largely attributable to infectious complications and organ failures.

Although more basic research is needed to better understand the pathogenesis of this entity, new treatment strategies that optimize currently available therapies are also needed to improve the outcome of cGVHD. Approximately 40% of patients do not respond to initial therapy.<sup>6,7</sup> Other studies show that some patients are overdiagnosed and overtreated.<sup>8</sup> Response rates to prednisone, azathioprine, cyclosporine, or combination regimens varied considerably among clinical trials,<sup>9,10</sup> implying patient selection plays an important role in treatment outcome.<sup>11</sup> Identifying patients most likely to have a poor outcome with cGVHD may allow intensification of prophylaxis and therapy for high-risk patients and avoid unnecessary immunosuppression in others. In the absence of a reliable stratification system, it is also difficult to conduct rigorous clinical trials. Determining prognostic factors for survival in patients with cGVHD would be a valuable tool in interpreting results of clinical trials with new agents.

However, the definition of risk status in patients with cGVHD is controversial. Although highly reproducible among transplantation centers,<sup>12</sup> the current grading system of limited versus extensive

From the Department of Oncology, Johns Hopkins University, Baltimore, MD; International Bone Marrow Transplantation Registry Graft-Versus-Host Disease Working Group, Milwaukee, WI; Fred Hutchinson Cancer Research Center, Seattle, WA; University of Nebraska Medical Center, Omaha; and University of Minnesota, Minneapolis.

Submitted October 17, 2002; accepted March 14, 2003. Prepublished online as *Blood* First Edition Paper, April 24, 2003; DOI 10.1182/blood-2002-10-3141.

Supported in part by Roche (Investigator-initiated Research Grant, CEL-151), Supergen (Unrestricted Research Grant), Sangstat (Unrestricted Research Grant), and K24CA83804 (G.B.V.). Presented in part in the Oral Session at the

43rd annual meeting of the American Society of Hematology, Orlando, FL, 2001.

**Reprints:** Georgia B. Vogelsang, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Bunting and Blaustein Cancer Research Bldg, 2M-89, 1650 Orleans St, Baltimore, MD 21231; e-mail: vogelge@jhmi.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2003 by The American Society of Hematology

cGVHD, originally proposed in 1980 based on the clinicopathologic findings in 20 patients,<sup>13</sup> has limitations. It only divides patients into those needing treatment (extensive cGVHD) and those who do not (limited cGVHD). A significant proportion of patients falls into the extensive cGVHD category and there is great heterogeneity in manifestations of cGVHD and patient outcomes within this group. Limitations of this classification system become more apparent as the incidence of cGVHD and the need for conducting randomized clinical trials increase.

Despite growing interest in developing models to predict the risk of dying in patients with cGVHD, it has been extremely difficult to devise a more helpful grading system. Several clinical features have had prognostic significance in previous studies, including "extensive" cGVHD<sup>9,13</sup> (multiorgan or extensive cutaneous involvement), Karnofsky performance status,<sup>9</sup> thrombocytopenia (< 100 000 cells/ $\mu$ L),<sup>9,10,14</sup> progressive-type onset,<sup>14,15</sup> lichenoid histology,<sup>15</sup> elevated bilirubin (> 1.2 mg/dL<sup>15</sup> or > 2.0 mg/dL<sup>16</sup>), age 20 or older,<sup>14</sup> gastrointestinal involvement,<sup>14</sup> no response to therapy at 6 months,<sup>14</sup> and subjective measures of severity.<sup>17</sup> Because of limitations in the numbers of subjects studied, heterogeneity of patient populations, and variations in diagnosis and management of cGVHD, no prognostic factors except for thrombocytopenia (TP) and progressive-type onset (PTO) have been reproduced by other investigators. Consequently, there has been no consensus about the best grading system for cGVHD. Risk stratification by platelet count has been validated by Przepiora and her colleagues.<sup>18</sup> In their data set, the long-term progression-free survival was 31% for patients without cGVHD, 51% for patients with low-risk (not thrombocytopenic), and 16% for patients with high-risk (thrombocytopenic) cGVHD.

We recently developed a prognostic model for cGVHD-specific survival by analyzing data on 151 consecutive patients who underwent allogeneic SCT at Johns Hopkins Oncology Center and subsequently had a diagnosis of cGVHD.<sup>19</sup> We studied the relationship between pretreatment clinical and laboratory features and cGVHD-specific survival (survival without recurrent malignancy). First, we examined these variables individually. Then we created a multivariable model using a stepwise modeling approach based on proportional hazards regression. We confirmed that TP and PTO were independent risk factors at diagnosis of cGVHD for shortened survival. We also identified extensive skin GVHD (ESI) involving more than 50% of body surface area (BSA) as another independent risk factor (RF) affecting the survival. A composite prognostic factor score (PS) was calculated for each patient by adding together the weighted averages of each of the RFs (weighted by the estimated regression coefficient). Then, patients were stratified into different prognostic groups based on their total scores. This prognostic model constituted the basis of the proposed new grading system in cGVHD.

This report describes the performance of this new prognostic scoring model and the individual RFs in 4 external data sets from different institutions.

## Patients, materials, and methods

Five comprehensive transplantation centers and the International Bone Marrow Transplant Registry (IBMTR) were asked to participate in this multicenter project in cGVHD. Two centers declined participation. University of Nebraska Medical Center (UN), IBMTR, Fred Hutchinson Cancer Research Center (FHCRC), and University of Minnesota (UM) contributed survival data from a total of 1105 cGVHD patients with complete clinical and follow-up information. All participating centers were asked to create a

separate data set in which the model could be tested independently. RF data for each patient at the time of initial diagnosis of cGVHD (or at the time of first reporting to the IBMTR) were considered. Date of death, the last follow-up date, and cause of death were included. Data were collected by all centers by reviewing the patient forms, abstracts, or medical records using standard procedures after obtaining an approval from the institutional review boards.

The extent of skin involvement in FHCRC patients was retrieved retrospectively from patients' abstracts using the "rule of nines" burn scale, which is similar to the method used at Hopkins. The data on skin GVHD were available in the other data sets, but the extent of involvement was either described differently (IBMTR) or was missing (UN and UM). In the IBMTR data collection forms, skin involvement had been originally coded as "mild," "moderate," and "severe." Therefore, these data were recoded as equal to or less than 50% BSA if it was originally recorded as "mild skin GVHD" and more than 50% if it was "moderate or severe skin GVHD." The effect of this classification system versus one in which only "severe" involvement was considered more than 50% was then tested in sensitivity analysis. In the UN and UM cohorts, the extent of skin involvement was considered more than 50% if cGVHD was originally classified as "extensive" cGVHD with skin involvement and equal to or less than 50% for those who had "limited" cGVHD with skin involvement.

Following the creation of the database, a prognostic score (PS) for each patient was calculated as described in the original study.<sup>19</sup> The follow-up data in the learning sample were updated as of April 2002 and regression coefficients (weighted averages) of the prognostic factors were recalculated before testing the model and the individual RFs in these external data sets. The coefficients by which they are multiplied were 1.949 for ESI, 1.293 for TP, and 0.514 for PTO. PS for any given patient was derived by adding together the products of the coefficient of each of the 3 significant RFs at the time of diagnosis and the value of the factor (0 if absent, 1 if present).

The formula for the calculation of PS is:  $PS = [1.949 \times (\text{skin extent})] + [1.293 \times (\text{platelets})] + [0.514 \times (\text{type of onset})]$ . The following conditions apply: (1) if the extent of skin involvement is more than 50% of BSA, put 1, otherwise put 0; (2) if the platelet count less than 100 000 cells/ $\mu$ L, put 1, otherwise put 0; and (3) if the cGVHD is PTO, put 1, otherwise put 0.

Then, Kaplan-Meier survival curves were plotted. Patients who died of relapsed malignancies or those who were alive at last follow-up were censored. The median follow-up for survivors and survival rates at 3 years together with the 95% confidence interval (CI) were reported. The hazard ratio (HR) and 95% CI comparing the prognostic groups were estimated using the Cox proportional hazards model. Multivariate models for ESI, TP, and PTO were performed to examine the effect of ESI on survival outcome after adjusting for TP and PTO. HRs and the corresponding 95% CIs were estimated using the Cox proportional hazards model. Analyses were performed cohort by cohort. A 2-sided  $P < .05$  from the Cox proportional hazards model was considered significant.

## Results

### Learning (training) sample

A total of 151 patients with a median age of 33 years (range, 4-62 years) had been included in the original study.<sup>19</sup> Patient characteristics are summarized in Table 1. Twenty-three patients (16%) were younger than 21 years old. A total of 140 patients (93%) received a stem cell transplant from an HLA-identical donor. Fifty-three percent of patients received their transplants before 1990. Onset of cGVHD was a median of 129 days (range, 27-1026 days) after SCT. Fifty-three patients (35%) had PTO cGVHD. Forty-five percent of patients had ESI. Platelet count was less than 100 000 cells/ $\mu$ L in 71 (47%) patients.

As of April 2002, the median follow-up for surviving patients was 8.3 years. The probability of cGVHD-specific survival at

Table 1. Patient characteristics

	Hopkins	Nebraska	IBMTR	FHCRC	Minnesota
n	151	60	708	188	149
Median age, y (range)	33 (4-62)	40 (17-60)	35 (16-60)	37 (1-70)	29 (1-52)
Age groups, n (%)					
Younger than 10 y	7 (5)	0 (0)	0	18 (10)	25 (17)
10-20 y	16 (11)	1 (2)	55 (8)	26 (14)	24 (16)
21-40 y	86 (60)	32 (53)	411 (58)	66 (35)	68 (46)
Older than 40 y	42 (28)	27 (45)	242 (34)	78 (41)	32 (21)
Male/female, n (%)	89 (59)/62 (41)	35 (58)/25 (42)	415 (59)/293 (41)	113 (60)/75 (40)	88 (59)/61 (41)
Donor-patient sex mismatch, n (%)					
None	76/151 (51)	36 (60)	358 (51)	86 (46)	71 (48)
Female to male	44/151 (29)	14 (23)	213 (30)	52 (28)	52 (35)
Male to female	30/151 (20)	10 (17)	137 (19)	50 (27)	26 (17)
Diagnosis, n (%)					
Aplastic anemia	4 (3)	0	0	2 (1)	16 (11)
Acute leukemia	59 (39)	17 (28)	304 (43)	63 (34)	40 (27)
CML	69 (46)	22 (37)	404 (57)	76 (40)	59 (39)
Lymphoma/CLL	13 (9)	10 (17)	0	20 (11)	4 (3)
Others	6 (4)	11 (18)	0	27 (14)	30 (20)
Disease status, n (%)					
Early (CR1 and chronic phase)	93 (62)	26 (43)	562 (79)	92 (60)	54 (36)
Late (all others)	58 (38)	34 (57)	146 (21)	61 (40)	95 (64)
Donor type, n (%)					
HLA-identical, related	140 (93)	60 (100)	708 (100)	51 (27)	96 (64)
HLA-identical, unrelated	7 (5)	0	0	117 (62)	18 (12)
Non-HLA-matched	4 (2)	0	0	20 (11)	35 (24)
Transplantation year, n (%)					
1979-1989	80 (53)	0	0	0 (0)	72 (48)
1990-1998	71 (47)	60 (100)	708 (100)	188 (100)	77 (52)
Conditioning regimen, n (%)					
Cy-TBI	71 (47)	12 (20)	708 (100)	139 (74)	98 (66)
Bu-Cy	62 (41)	1 (2)	0	29 (15)	10 (7)
Bu-Cy-VP16 or others	18 (12)	47 (78)	0	20 (11)	41 (27)
Source of stem cells, n (%)					
Bone marrow	151 (100)	39 (65)	708 (100)	157 (84)	149 (100)
Blood	0	21 (35)	0	31 (16)	0
GVHD prophylaxis, n (%)					
CSA ± MTX	141 (93)	60 (100)	708 (100)	156 (83)	33 (22)
FK506 + MTX		0	0	7 (4)	9 (6)
Others	10 (7)	0	0	25 (13)	100 (68)
None	0	0	0	0	2 (1)
T-cell depletion, n (%)	33 (22)	0	0	6 (3)	5 (3)
Positive CMV status, n (%)	79/138 (57)	25 (42)	555† (78)	70 (38)	100 (67)
ABO incompatibility, n (%)	43/143 (30)	N/A	N/A	N/A	N/A
Acute GVHD stage, n (%)					
0	31/151 (21)	15 (25)	247 (35)	4 (3)	19 (13)
1 (cutaneous)	65/151 (43)	12 (20)	198 (28)	3 (2)	18 (12)
2-4 (systemic)	55/151 (36)	33 (55)	263 (37)	136 (95)	112 (75)
Follow-up,* y after diagnosis of cGVHD, median (range)	8.3 (0.2-20.6)	4.3 (0.5-7.5)	1.2 (0.1-5.2)	1.6 (0.1-5.1)	8.4 (0.7-13.4)
Mode of presentation of cGVHD, n (%)					
Progressive	53 (35)	16 (27)	156 (22)	53 (28)	69 (46)
Quiescent	67 (44)	32 (53)	306 (43)	115 (61)	62 (42)
De novo	31 (21)	12 (20)	246 (35)	20 (11)	18 (12)
Skin extent, n (%)					
None	32 (21)	40 (66)	209 (30)‡	76 (40)	34 (23)
50% or less	52 (34)	5 (9)	318 (45)‡	70 (37)	7 (5)
More than 50%	67 (45)	15 (25)§	181 (26)‡	42 (22)	108 (72)§
Weight loss, n (%)					
None	41/150 (27)	N/A	546 (77)	N/A	N/A
Less than 10% from baseline (BMT)	50/150 (33)	N/A	N/A	N/A	N/A
10% or more from baseline (BMT)	59/150 (40)	N/A	N/A	N/A	N/A

**Table 1. Patient characteristics (continued)**

	Hopkins	Nebraska	IBMTR	FHCRC	Minnesota
<b>Karnofsky performance status, n (%)</b>					
Greater than 80%	84 (56)	N/A	303 (43)	74 (64)	N/A
50%-80%	57 (38)	N/A	385 (54)	36 (31)	N/A
Less than 50%	10 (6)	N/A	20 (3)	5 (4)	N/A
<b>Mouth/eye involvement, n (%)</b>					
Asymptomatic	31 (20)	N/A	233 (33)	N/A	101 (68) <sup>  </sup>
Symptomatic	99 (66)	N/A	Either 317 (45)	N/A	48 (32) <sup>  </sup>
Functional compromise	21 (14)	N/A	Both 158 (22)	N/A	
<b>Thrombocytopenia less than 100 000/<math>\mu</math>L, n (%)</b>	71 (47)	26 (43)	265 (37)	129 (69)	88 (59)

Except where otherwise indicated all data are numbers of patients, with percentages in parentheses.

CML indicates chronic myeloid leukemia; CLL, chronic lymphocytic leukemia; CR, complete remission; Cy, cytarabine; TBI, total body irradiation; Bu, busulfan; CMV, cytomegalovirus; BMT, bone marrow transplantation; N/A, not available.

\*Of surviving patients.

†Either patient or donor CMV<sup>+</sup> or unknown.

‡Extent of skin involvement was considered greater than 50% BSA if skin GVHD was originally recorded as “moderate” or “severe” in the IBMTR data collection forms.

§Extent of skin involvement was considered greater than 50% BSA if cGVHD was classified as “extensive” with skin involvement.

<sup>||</sup>In the Minnesota study, mouth/eye involvement was considered “present” (symptomatic) or “absent” (asymptomatic).

3 years after the diagnosis of cGVHD was 58% (95% CI, 49%, 65%; (Figure 1A). At the time of analysis, 67 patients (44%) had died from causes other than relapse of underlying hematologic malignancies. Fifty-eight patients (87%) died from various infectious complications during the course of cGVHD. Seventeen of 151 patients (11%) died from relapse of underlying malignancies.

According to univariate analysis, extensive skin GVHD involving more than 50% of BSA (ESI), poor performance status (Karnofsky performance status < 80%), TP (< 100 000 cells/ $\mu$ L), PTO, date of transplantation, low absolute serum immunoglobulin G level (< 500 mg/dL), hyperbilirubinemia (> 1.2 mg/dL), and presentation with a systemic infection at the diagnosis of cGVHD were significant RFs for shortened survival. Multivariable analysis demonstrated that ESI, TP, and PTO were independently associated with shortened survival with respective HRs of 7.0, 3.6, and 1.7. The probability of survival at 3 years for 54 patients with PS 0 (favorable-risk group; 0 RF) was 92% (95% CI, 80%, 97%). Forty-seven patients with PS between 0 and 2 (intermediate-risk group; 1 RF) had 71% (95% CI, 55%, 82%) and 50 patients with PS equal to or greater than 2 (high-risk group; > 1 RF) had 9% (95% CI, 3%, 19%) probability of survival at 3 years (Figure 1B). The HR for mortality of the intermediate-risk group compared with the favorable-risk group was 3.7 (1.4-9.3; *P* = .007); the HR for the high-risk group compared with the intermediate-risk group was 6.9 (3.8-12.4; *P* < .001).

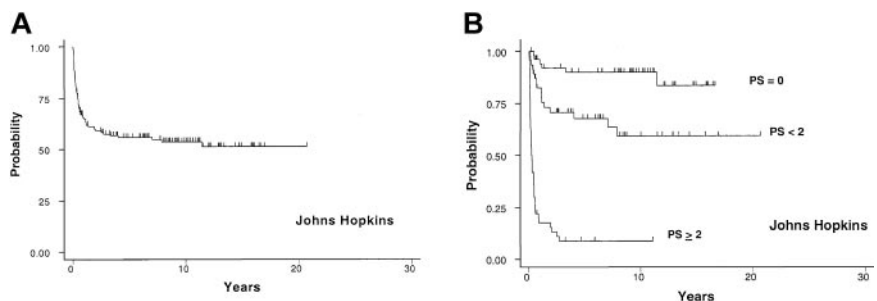
**Test samples**

Baseline clinical and laboratory characteristics at the time of diagnosis of cGVHD of the learning and test samples are summarized in Table 1. The distribution of the 3 independent RFs in test samples and the number of patients in each of these 3 prognostic

subgroups are summarized in Table 2. The distinct clinical features and predictive power of the model and individual RFs in each of these test samples are summarized below.

**Cohort 1.** The first test sample included 60 patients with a median age of 40 years (range, 17-60 years) from the UN. Patients developed cGVHD following HLA-matched related donor SCT performed between 1990 and 1996. Twenty-one patients (35%) received blood stem cell transplants. Acute GVHD prophylaxis was with cyclosporine (CSA) and methotrexate (MTX) for all patients. cGVHD was confirmed by histology and appeared a median of 192 days after the transplantation. The median follow-up after diagnosis of cGVHD for surviving patients was 4.3 years. The 3-year probability of cGVHD-specific survival was 52% with 95% CI, 38% to 64% (Figure 2A). The mortality HR for comparison of the intermediate-risk to the favorable-risk group was 2.3 and the HR for comparison of the high-risk to the intermediate-risk group was 1.6 (Table 3). The 3-year probabilities of cGVHD-specific survival for 23 (38%), 30 (50%), and 7 (12%) patients in the favorable-, intermediate-, and high-risk groups were 77% (95% CI, 53%, 90%), 42% (95% CI, 24%, 59%), and 14% (95% CI, 0.7%, 46%), respectively (Figure 2B). Of note, 2 patients in the favorable-risk group died at 6 and 6.5 years without evidence of active cGVHD. TP was the only significant predictor of survival after adjusting for the other 2 RFs (Table 4).

**Cohort 2.** The IBMTR data included 708 patients with a median age of 35 years (range, 16-60 years) who underwent allogeneic bone marrow transplantation from an HLA-identical sibling donor between 1990 and 1994 at 119 centers with data on 1 to 52 patients per center. cGVHD was diagnosed in a median of 134 days (range, 60-1025 days) after transplantation. Twenty-two percent of patients presented with progressive-type cGVHD.



**Figure 1. Johns Hopkins cohort.** Probability of cGVHD-specific survival in years after the diagnosis of cGVHD in 151 allogeneic BMT recipients from Johns Hopkins (A) and the survival based on PS (B). The “PS < 2” label in this and other graphs indicates PS between 0 and 2 (> 0 and < 2).



**Table 2. Distribution of the risk factors in each cohort**

	N (%)				
	Hopkins	Nebraska	IBMTR	FHCRC	Minnesota
n	151	60	708	188	149
Distribution of RFs					
ESI alone	17 (11)	8 (13)	69 (10)	4 (2)	22 (15)
TP alone	21 (14)	10 (17)	117 (17)	74 (39)	12 (8)
PTO alone	4 (3)	2 (3)	45 (6)	6 (3)	5 (3)
ESI + TP	6 (4)	3 (5)	52 (8)	15 (8)	29 (20)
ESI + PTO	5 (3)	1 (2)	15 (2)	7 (4)	17 (11)
TP + PTO	5 (3)	10 (17)	51 (7)	24 (13)	7 (5)
ESI + TP + PTO	39 (26)	3 (5)	45 (6)	16 (8)	40 (27)
None	54 (36)	23 (38)	314 (44)	42 (23)	17 (11)
Prognostic score, no. of RFs					
0 (0 RF)					
Fewer than 2 (TP + PTO or only 1 RF)	54 (36)	23 (38)	314 (44)	42 (23)	17 (11)
2 or more (> 1 RF)	47 (31)	30 (50)	282 (40)	108 (57)	47 (32)
	50 (33)	7 (12)	112 (16)	38 (20)	85 (57)

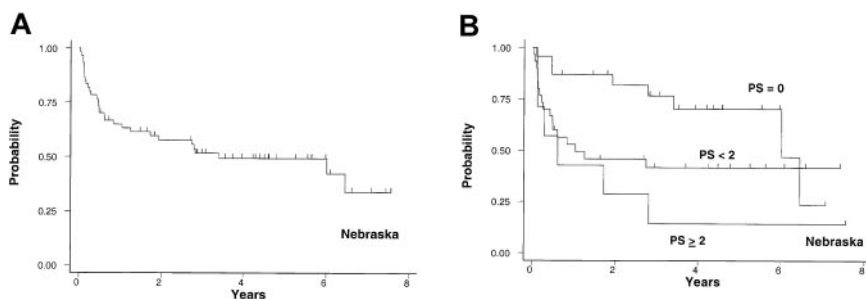
Twenty-six percent of patients had extensive skin GVHD and 37% had TP at the time of diagnosis of cGVHD or initial reporting to the IBMTR. Median follow-up for surviving patients after diagnosis of cGVHD was 1.2 years and the 3-year probability of cGVHD-specific survival for the entire group was 77% (95% CI, 72%, 81%; Figure 3A). The HR for comparison of the intermediate-risk to the favorable-risk group was 3.3 and the HR for comparison of the high-risk to the intermediate-risk group was 2.5 (Table 3). The 3-year probabilities of cGVHD-specific survival for 314 (44%), 282 (40%), and 112 (16%) patients in favorable-, intermediate-, and high-risk groups were 91% (95% CI, 85%, 95%), 74% (95% CI, 66%, 80%), and 49% (95% CI, 37%, 60%), respectively (Figure 3B). In this cohort, ESI was the strongest predictor of survival after adjusting for the other 2 RFs. The HR comparison of ESI (moderate/severe skin GVHD, coded as > 50% BSA) to non-ESI (mild/no skin GVHD, coded as ≤ 50% BSA) was 3.3 after adjusting for TP and PTO (Table 4). TP and PTO were also significant RFs for survival in this cohort.

The analysis was then repeated reclassifying only "severe" skin GVHD as ESI. However, this reduced the proportion of the IBMTR population with ESI to 7% (compared with 22%-72% in the other cohorts). ESI still remained an independent predictor for poor survival after adjusting for the other 2 RFs. The HR comparison of ESI (severe skin GVHD, coded as > 50% BSA) to non-ESI (none, mild and moderate skin GVHD, coded as < 50% BSA) was 2.0 after adjusting for TP and PTO ( $P = .01$ ).

**Cohort 3.** This data set included only patients diagnosed with extensive cGVHD at the FHCRC at day + 80 after transplantation cGVHD screening. Patients with extensive cGVHD diagnosed outside the FHCRC by their primary physicians were excluded because the information recorded for the extent of skin involvement was not provided in a consistent manner. A total of 188

patients who received transplants in 1995 or later were examined. The proportion of patients who underwent HLA-matched unrelated donor transplantation was higher than those who had matched related and mismatched related transplantations combined. Bone marrow was the source of stem cells in 84% of the cases. Of 143 graded cases, 136 (95%) developed grade 2 to 4 acute GVHD after transplantation. Forty-two patients (22%) had ESI, 53 (28%) presented with PTO of cGVHD, and 129 (69%) of patients had TP at the time of diagnosis of cGVHD. Relatively few patients in this cohort of 188 patients had ESI without PTO (10%) or PTO without ESI (16%). Also only 16 of 188 patients (8%) had all 3 RFs at diagnosis of cGVHD (Table 2). Median follow-up for 157 surviving patients after diagnosis of cGVHD was 1.6 years, ranging from 0.1 to 5.1 years. The 3-year probability of cGVHD-specific survival for the entire group was 77% (95% CI, 69%, 84%; Figure 4A). The 3-year probabilities of cGVHD-specific survival for 42 (22%), 108 (58%), and 38 (20%) patients in favorable-, intermediate-, and high-risk groups were 96% (95% CI, 76%, 99%), 77% (95% CI, 65%, 85%), and 62% (95% CI, 42%, 77%), respectively (Figure 3B). The HR for comparison of the intermediate-risk group to the favorable-risk group was 8.9 and the HR for comparison of the high-risk group to the intermediate-risk group was 1.6 (Table 3). TP was the only significant predictor of survival after adjusting for the other 2 RFs (Table 4).

**Cohort 4.** This cohort consisted of 149 patients with a median age of 29 years (range, 1-52 years) in whom the diagnosis of cGVHD after allogeneic SCT was made at the UM. Forty-eight percent of patients received their transplants before 1990. This cohort contained the highest proportion of pediatric patients. Twelve percent of patients had an HLA-matched unrelated donor and 24% underwent non-HLA-matched allogeneic SCT. Of 100 patients who received GVHD prophylaxis other than the standard



**Figure 2. University of Nebraska cohort.** Probability of cGVHD-specific survival in years after the diagnosis of cGVHD in 60 patients from the University of Nebraska (A) and the survival based on PS (B).

**Table 3. Comparative HRs for survival in each participating center**

	Intermediate versus favorable risk			High versus intermediate risk		
	HR	95% CI	P	HR	95% CI	P
Hopkins	3.7	1.4, 9.3	.007	6.9	3.8, 12.4	< .0001
Nebraska	2.3	1.0, 5.3	.06	1.6	0.6, 4.0	.34
IBMTR	3.3	1.9, 5.7	< .001	2.5	1.7, 3.7	< .001
FHCRC	8.9	1.2, 66.9	.03	1.6	0.8, 3.3	.22
Minnesota	2.6	0.8, 8.8	.12	1.6	1.0, 2.7	.07

P values are 2-sided.

CSA (or FK-506) and MTX, 63 of these received MTX, antithymocyte globulin, and prednisone, 22 received CSA and prednisone, and 15 received CSA, MTX, and prednisone. The median time from transplantation to onset of cGVHD was 141 days (range, 52-775 days). Forty-six percent had progressive onset and 72% had extensive skin cGVHD. Fifty-nine percent of patients had TP. This cohort had the highest percentage (63%) of patients with at least 2 RFs at diagnosis of cGVHD (Table 2). For the entire group, the probability of cGVHD-specific survival at 3 years after the diagnosis of cGVHD was 54%, (95% CI, 45%, 62%; Figure 5A). The 3-year probabilities of survival for 17 (11%), 47 (32%), and 85 (57%) patients in favorable-, intermediate- and high-risk groups were 81% (95% CI, 51%, 93%), 68% (95% CI, 52%, 80%), and 41% (95% CI, 30%, 52%), respectively (Figure 5B). The HR for comparison of the intermediate-risk to the favorable-risk group was 2.6 and the comparison of the high-risk to the intermediate-risk group was 1.6 (Table 3). TP and PTO were independent predictors for survival after adjusting for the other RFs (Table 4).

## Discussion

As in any other disease, one of the critical considerations in making a therapeutic decision for a patient with cGVHD is predicting the likely outcome. Accurate clinical grading or classification schemes may allow us to intelligently decide between earlier or more intensive immunosuppressive therapy for patients considered high-risk for cGVHD-related death versus less intensive approaches for those intrinsically destined to have a more favorable outcome. Therefore, classification of patients with cGVHD based on their prognosis is clinically more relevant than purely anatomic classification into limited versus extensive disease.

We previously identified 3 RFs that are predictive for chronic GVHD-specific survival.<sup>19</sup> Excluding relapse-related deaths, 4-level<sup>19</sup> or 3-level<sup>20</sup> prognostic models based on these risk factors predicted the probability of being alive at 10 years ranging from 9% to 90%. The difference in this cGVHD-specific survival at 10 years was 30% between favorable- and intermediate-risk groups and 50% between intermediate- and high-risk groups.<sup>20</sup> The data set included many patients receiving transplants before 1990 and, as of April 2002, median follow-up was over 8 years. The primary assertion of the initial study

was that consideration of skin involvement added significant prognostic information for predicting nonrelapse mortality among patients with newly diagnosed cGVHD.

Using multiple data sets that included a total of 1105 patients, we tested the validity of (1) a prognostic scoring model and (2) each of the 3 prognostic factors that constituted this model. These cohorts represented heterogeneous patient populations with different sample sizes and distinct clinical characteristics that were treated in different institutions. The median survivals and the follow-ups, whether or not divided into prognostic subgroups, were also not comparable among the cohorts. Differences in the date of transplantation and date of diagnosis of cGVHD seem likely to be the major cause for discrepancies in survival outcome among these cohorts. Many of the patients in the learning sample were diagnosed with cGVHD before day + 100, and some patients underwent transplantation in an era before the use of CSA.

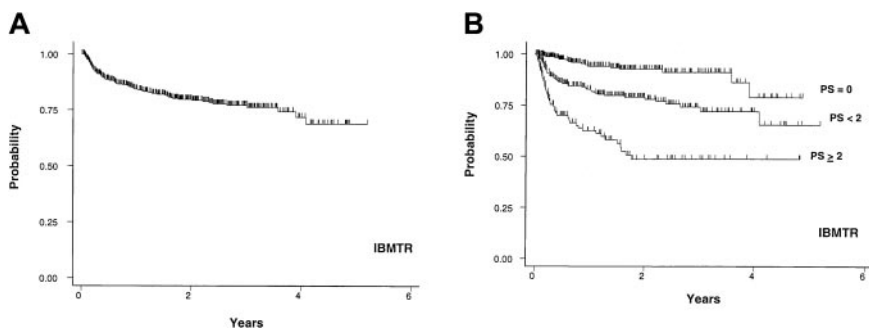
The distribution of the RFs among these groups was also not comparable possibly because of the variability in the method and timing of diagnosis of cGVHD as well as differences in the diagnostic criteria used for acute and chronic GVHD. One of the most difficult aspects of the management of cGVHD is making a timely diagnosis. Our previous data clearly indicate that making an accurate and early diagnosis is a critical step in the management of cGVHD.<sup>8</sup> In the learning sample, all but 2 patients had their cGVHD confirmed by biopsy. As mentioned in "Materials and methods," the FHCRC cohort was limited to patients diagnosed with cGVHD at the day + 80 evaluation because of the availability of on-site skin evaluation. Patients with early-onset cGVHD and those who had the diagnosis of extensive cGVHD outside this transplantation center were not included. This approach might have selected some of the patients with different clinical characteristics and prognoses.

Therefore, our conclusions may be limited by the use of data sets with heterogeneous populations, because no study has shown the RFs would be the same for cGVHD in children versus adults, T cell-depleted versus non-T cell-depleted recipients, and so forth. However, the ability of the prognostic model to separate groups with such variable survival suggests the potential utility of the proposed grading schema after prospective validation. Despite the heterogeneity, the separation of the survival curves based on the PS was quite good in all cohorts. In contrast to the learning sample, the magnitude of difference between the HRs of favorable-risk and

**Table 4. Multivariate models for each of the RFs in all centers**

	Skin GVHD more than 50% BSA			Platelet count less than 100 000/ $\mu$ L			PTO		
	HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Hopkins	7.0	3.7, 13.5	< .001	3.6	1.9, 6.8	< .001	1.7	0.9, 3.0	.08
Nebraska	1.2	0.5, 2.6	.7	3.2	1.4, 7.3	.005	1.3	0.6, 2.9	.5
IBMTR	3.3	2.2, 4.8	< .001	1.8	1.2, 2.7	.004	1.5	1.0, 2.3	.03
FHCRC	1.4	0.7, 3.2	.4	3.3	1.1, 9.3	.03	1.9	0.8, 4.1	.1
Minnesota	1.2	0.7, 2.1	.6	2.5	1.5, 4.2	.001	1.7	1.1, 2.8	.03

P values are 2-sided.



**Figure 3. IBMTR cohort.** Probability of cGVHD-specific survival in years after the diagnosis or first reporting of cGVHD for 708 patients from the IBMTR (A) and the survival based on PS (B).

intermediate-risk groups was greater than the difference between the HRs of intermediate- and high-risk groups in all test samples (Table 3). In the FHCRC cohort, although the HRs comparing the favorable-risk group with the 2 other groups were statistically significant, the difference between intermediate- and high-risk groups did not reach the conventional value of statistical significance despite different survival likelihood. This might be due to limited power from the small number of high-risk patients.

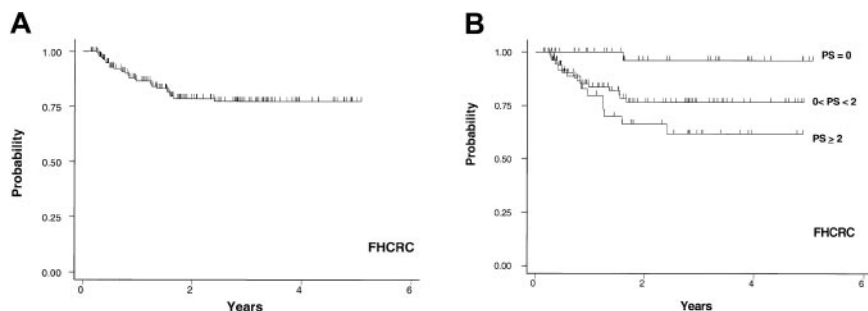
In the original paper, the major new finding was the extremely high mortality hazard (7.0) associated with skin GVHD involving more than 50% BSA. None of the test cohorts showed a similarly high mortality hazard associated with skin involvement more than 50% BSA, and 3 of the 4 test cohorts showed no statistically significant association between the extent of skin involvement at the time of diagnosis and subsequent mortality after adjusting for TP and PTO. Reasons for the strikingly different HR estimates among different cohorts are not clear but might be related to selection criteria or methods for collecting this information. The generally similar HR estimates associated with TP and PTO suggest that methods for measuring and reporting the extent of skin involvement were more important than patient selection in accounting for the lack of uniformity in HR estimates associated with skin involvement.

A possible limitation of the study was the absence of quantitative information about the extent of cGVHD skin involvement in the test samples except for cohort 3 (FHCRC). To estimate this information, the available data on skin GVHD was quantified using a predetermined criterion for each of these cohorts. This approach might have underestimated or overestimated the actual extent of skin GVHD in some of these patients. To evaluate the effect of this bias, we applied the model in the same IBMTR data set after setting a cutoff between “moderate” and “severe” skin GVHD. When it was reclassified as having only “severe” skin GVHD, ESI remained a significant predictor of cGVHD-specific survival. Excluding patients who were originally coded as “moderate” skin GVHD from the analysis to eliminate the possible misclassification bias resulted in similar results. Thus, it is possible but not likely that our interpretation of ESI based on the IBMTR categories of “mild, moderate, and severe” influenced conclusions in this data set. Regardless, the validity of the conclusion about the prognostic

implication of the extensive skin GVHD may be limited by the contrived method of assessing skin involvement in data sets where this was not collected prospectively. Therefore, a prospective study using standardized criteria for organ involvement is needed to confirm the utility of the skin extent in this prognostic model.

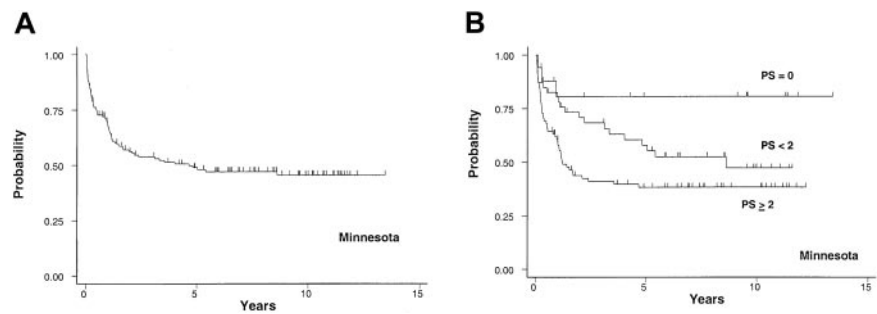
The other explanation for the variable HR estimates associated with prognostic model in these data sets may be related to the study end point we used. As seen in the survival curves of UN, 2 patients in the favorable-risk group died at 6 and 6.5 years after cGVHD because of late events not related to GVHD or relapse. cGVHD is a unique syndrome. Relapse of underlying malignancy is the other major competing factor (10%) for death in patients with chronic GVHD.<sup>14,18,19</sup> We also observed that the probability of relapse was inversely correlated with the occurrence and severity<sup>19</sup> of cGVHD. We did not stratify patients based on entry diagnosis and disease status in our original design. Therefore, we have used cGVHD-specific survival (survival without recurrent malignancy) as the study end point in the current studies. Overall survival would have been a valid and clinically relevant end point if there were sufficient patients in each disease category and disease status in comparing cGVHD. It is possible that for each disease and remission status a different combination of factors would be found, which may be interesting, but would provide a very difficult model to use clinically in managing patients. Certainly, we need other end points when considering prospective validation of the model and the RFs. Measurements of morbidity and time to discontinuation of immunosuppressive treatment are potentially valuable end points, which should be included in future studies in cGVHD.

Despite its limitations, the new clinical grading or stratification system based on the PS model can be useful to individualize treatment plans in cGVHD. In a recent randomized study comparing single-agent steroid with CSA and steroid, the hazards of transplantation-related mortality, overall mortality, recurrent malignancy, secondary therapy, and discontinuation of all immunosuppressive therapy were not significantly different between the 2 arms, but survival without recurrent malignancy was lower in the 2-drug arm.<sup>21</sup> In light of these new findings, it could be even more appropriate to treat patients in the favorable-risk category with



**Figure 4. FHCRC cohort.** Probability of cGVHD-specific survival in years after the diagnosis of cGVHD in 188 patients from the Fred Hutchinson Cancer Research Center (A) and the survival based on PS (B).

**Figure 5. University of Minnesota cohort.** Probability of cGVHD-specific survival in years after the diagnosis of cGVHD in 149 patients from the University of Minnesota (A) and the survival based on PS (B).



corticosteroids alone. Alternatively, new treatment options other than corticosteroids may be sought in this group of patients. If patients fail to respond to steroids or other single agents, combination therapies might then be instituted. In contrast, patients with limited life expectancy due to poor-risk cGVHD may benefit from new therapeutic approaches. Clearly, the cGVHD-related mortality in this group is unacceptably high with current treatment approaches and new treatment paradigms are needed.

In future prospective trials evaluating any new approaches to cGVHD therapy, patients should be stratified according to the new proposed clinical grading to ensure a balanced distribution of RFs among groups. To improve comparability between publications, reports of cGVHD treatment trials should include an accurate

description of the study population based on characteristics of the disease at the time of diagnosis. In view of the limitations of the current grading system, this new clinical grading system could also be used to report severity of cGVHD instead of (or in addition to) the limited/extensive stage.

In conclusion, the prognostic model including the 3 RFs (ESI, TP, and PTO) is promising for clinical application because of its utility across multiple independent data sets. The utility of skin involvement must be validated prospectively before routinely including this characteristic as an RF for mortality. Until then, this clinical grading system could be used in its present form to improve clinical management by adjusting therapy, to group patients for clinical trials, and to ensure better communication among the centers.

## References

- Sullivan KM, Agura E, Anasetti C, et al. Chronic graft-versus-host disease and other late complications of bone marrow transplantation. *Semin Hematol.* 1991;28:250-259.
- Marmont AM, Horowitz MM, Gale RP, et al. T-cell depletion of HLA-identical transplants in leukemia. *Blood.* 1991;78:2120-2130.
- Cutler C, Giri S, Jeyapalan S, et al. Acute and chronic graft-versus-host disease after allogeneic peripheral-blood stem-cell and bone marrow transplantation: a meta-analysis. *J Clin Oncol.* 2001;19:3685-3691.
- Akpek G, Boitnott JK, Lee LA, et al. Hepatic-variant of graft-versus-host disease after donor lymphocyte infusion. *Blood.* 2002;100:3903-3907.
- Socie G, Stone JV, Wingard JR, et al. Long-term survival and late deaths after allogeneic bone marrow transplantation. Late Effects Working Committee of the International Bone Marrow Transplant Registry. *N Engl J Med.* 1999;341:14-21.
- Sullivan KM, Witherspoon R, Storb R, et al. Alternating-day cyclosporine and prednisone for treatment of high-risk chronic graft-versus-host disease. *Blood.* 1988;72:555-561.
- Sullivan KM, Witherspoon RP. Cyclosporine treatment of chronic graft-versus-host disease following allogeneic bone marrow transplantation. *Transplantation.* 1990;22:1336-1338.
- Jacobsohn DA, Montross S, Anders V, Vogelsang GB. Clinical importance of confirming or excluding the diagnosis of chronic graft-versus-host disease. *Bone Marrow Transplant.* 2001;28:1047-1051.
- Sullivan KM, Shulman HM, Storb R, et al. Chronic graft-versus-host disease in 52 patients: adverse natural course and successful treatment with combination immunosuppression. *Blood.* 1981;57:267-276.
- Sullivan KM, Witherspoon RP, Storb R, et al. Prednisone and azathioprine compared with prednisone and placebo for treatment of chronic graft-versus-host disease: prognostic influence of prolonged thrombocytopenia after allogeneic marrow transplantation. *Blood.* 1988;72:546-554.
- Lee SJ, Vogelsang G, Gilman A, et al. A survey of diagnosis, management, and grading of chronic GVHD. *Biol Blood Marrow Transplant.* 2002;8:32-39.
- Atkinson K, Horowitz MM, Gale RP, et al, for the Advisory Committee of the International Bone Marrow Transplant Registry. Consensus among bone marrow transplanters for diagnosis, grading and treatment of chronic graft-versus-host disease. *Bone Marrow Transplant.* 1989;4:247-254.
- Shulman HM, Sullivan KM, Weiden PL, et al. Chronic graft-versus host syndrome in man: a long-term clinicopathologic study of 20 Seattle patients. *Am J Med.* 1980;69:204-217.
- Arora M, Burns LJ, Davies SM, et al. Chronic graft versus host disease: a prospective cohort study [abstract]. *Blood.* 2001;98:1667a.
- Wingard JR, Piantadosi S, Vogelsang GB, et al. Predictors of death from chronic graft-versus-host disease after bone marrow transplantation. *Blood.* 1989;74:1428-1435.
- Pavletic S, Tarantolo S, Lynch J, et al. Chronic graft-versus-host disease after allogeneic blood stem cell or bone marrow transplantation: factors determining the onset and survival [abstract]. *Proc Am Soc Clin Oncol.* 1999;18:201a.
- Lee SJ, Klein JP, Barret AJ, et al. Severity of chronic graft-versus-host disease: association with treatment-related mortality and relapse. *Blood.* 2002;100:406-414.
- Przepiorka D, Anderlini P, Saliba R, et al. Chronic graft-versus-host disease after allogeneic blood stem cell transplantation. *Blood.* 2001;98:1695-1700.
- Akpek G, Zahurak ML, Piantadosi S, et al. Development of a prognostic model for grading chronic graft-versus-host disease. *Blood.* 2001;97:1219-1226.
- Akpek G. Clinical grading in chronic graft-versus-host disease: is it time for change? *Leuk Lymphoma.* 2002;43:1211-1220.
- Koc S, Leisenring W, Flowers ME, et al. Therapy for chronic graft-versus-host disease: a randomized trial comparing cyclosporine plus prednisone versus prednisone alone. *Blood.* 2002;100:48-51.