

Global and organ-specific chronic graft-versus-host disease severity according to the 2005 NIH Consensus Criteria

Sally Arai,¹ Madan Jagasia,² Barry Storer,³ Xiaoyu Chai,³ Joseph Pidala,⁴ Corey Cutler,⁵ Mukta Arora,⁶ Daniel J. Weisdorf,⁶ Mary E. D. Flowers,³ Paul J. Martin,³ Jeanne Palmer,⁷ David Jacobsohn,⁸ Steven Z. Pavletic,⁹ Georgia B. Vogelsang,¹⁰ and Stephanie J. Lee³

¹Division of Blood and Marrow Transplantation, Stanford University Medical Center, Stanford, CA; ²Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; ³Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA; ⁴Blood and Marrow Transplantation, Moffitt Cancer Center, Tampa, FL; ⁵Hematologic Malignancies, Dana-Farber Cancer Institute, Boston, MA; ⁶Blood and Marrow Transplant Program, University of Minnesota, Minneapolis, MN; ⁷Division of Hematology/Oncology, Medical College of Wisconsin, Milwaukee, WI; ⁸Division of Blood and Marrow Transplantation, Children's National Medical Center, Washington, DC; ⁹National Cancer Institute, National Institutes of Health, Bethesda, MD; and ¹⁰Department of Oncology, Johns Hopkins Hospital, Baltimore, MD

In 2005, the National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic GVHD proposed a new scoring system for individual organs and an algorithm for calculating global severity (mild, moderate, severe). The Chronic GVHD Consortium was established to test these new criteria. This report includes the first 298 adult patients enrolled at 5 centers of the Consortium. Patients were assessed every 3-6 months using standardized forms rec-

ommended by the Consensus Conference. At the time of study enrollment, global chronic GVHD severity was mild in 10% (n = 32), moderate in 59% (n = 175), and severe in 31% (n = 91). Skin, lung, or eye scores determined the global severity score in the majority of cases, with the other 5 organs determining 16% of the global severity scores. Conventional risk factors predictive for onset of chronic GVHD and nonrelapse mortality in people with chronic GVHD were not associated

with NIH global severity scores. Global severity scores at enrollment were associated with nonrelapse mortality ($P < .0001$) and survival ($P < .0001$); 2-year overall survival was 62% (severe), 86% (moderate), and 97% (mild). Patients with mild chronic GVHD have a good prognosis, while patients with severe chronic GVHD have a poor prognosis. This study was registered at www.clinicaltrials.gov as no. NCT00637689. (*Blood*. 2011;118(15): 4242-4249)

Introduction

Chronic GVHD is a common complication associated with high morbidity and mortality after allogeneic hematopoietic cell transplantation (HCT).^{1,2} Before 2005, chronic GVHD was only diagnosed after day 100, and the severity was described as “limited” (< 50% body surface area skin involvement or liver involvement only) or “extensive” (involvement of any other target organ, > 50% body surface area involvement or cirrhosis).³ In 2005, the National Institutes of Health (NIH) Consensus Working Group for Diagnosis and Staging recommended organ-specific severity scoring scales and proposed a new definition for global chronic GVHD severity.⁴ The new global severity score was intended to replace the “limited” versus “extensive” designation with the goal of providing a more clinically informative and discriminating severity measure for use in clinical trials and as an indicator of the need for systemic immunosuppressive treatment. The hope was also that better phenotypic classifications would assist laboratory researchers studying biologic correlates and pathophysiology of chronic GVHD.⁵ The NIH global severity score uses the numerical scoring system for individual organs to calculate a summary scale according to the number and severity of organs involved.

The Chronic GVHD Consortium is an NIH-funded study group established to test the new chronic GVHD criteria because the scoring system was based on consensus opinion and not empiric

data. Using data from a prospective observational cohort study, we sought to: (1) describe the global and organ-specific severity of participants; (2) describe the individual organ contributions within the global categories and ascertain whether all organs need to be scored; (3) assess clinical risk factors for higher global severity; and (4) assess whether global severity predicts nonrelapse mortality and overall survival.

Methods

The Chronic GVHD Consortium began patient accrual in 2007 and this report includes the first 298 adult patients from 5 centers: Fred Hutchinson Cancer Research Center, Stanford University, University of Minnesota, Dana-Farber Cancer Institute, and Vanderbilt University. The protocol was approved by the institutional review boards of participating centers, and all patients provided written informed consent in accordance with the Declaration of Helsinki. Eligible patients were HCT recipients age 2 or older with chronic GVHD, diagnosed according to the NIH consensus criteria, and requiring systemic immunosuppressive therapy. Patients with either classic chronic GVHD (without features of acute GVHD) or overlap syndrome (features of both chronic and acute GVHD) were eligible. Cases were classified as incident (enrollment < 3 months after chronic GVHD diagnosis) or prevalent (enrollment 3 or more months after chronic GVHD

Submitted March 24, 2011; accepted July 5, 2011. Prepublished online as *Blood* First Edition paper, July 26, 2011; DOI 10.1182/blood-2011-03-344390.

An Inside *Blood* analysis of this article appears at the front of this issue.

Presented in abstract form at the 52nd annual meeting of the American Society

of Hematology, Orlando, FL, December 6, 2010.

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 USC section 1734.

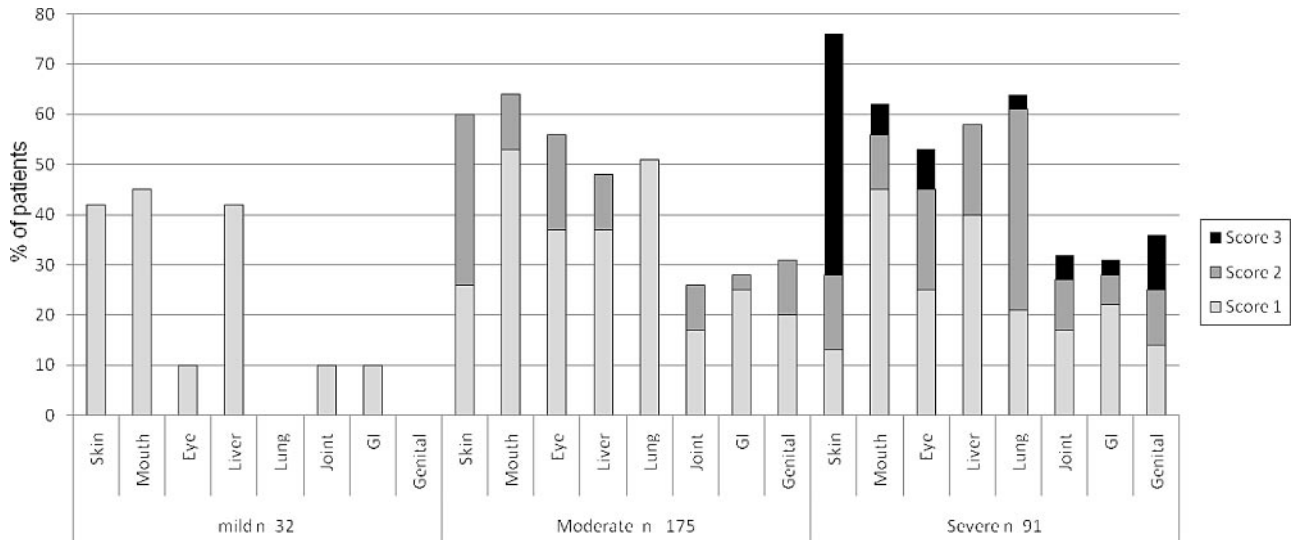


Figure 1. Individual organ severity scoring within global severity categories.

diagnosis). At enrollment and every 6 months thereafter, standardized data were collected from clinicians and patients as recommended by the NIH Consensus Conference.^{4,6} Incident cases had an additional assessment at 3 months after enrollment. All patients were followed with serial assessments until their chronic GVHD was resolved for 1 year. Collection of longitudinal data is ongoing. Data collection forms, the ACCESS database structure, and SAS coding programs are available on request from the authors.

Clinician organ severity scoring

A clinical categorical system (0-3) is used for scoring of individual organs that describes the severity for each affected organ taking functional impact into account.⁴ Eight organs (skin, mouth, eyes, gastrointestinal [GI] tract, liver, lungs, joints, and female genital tract) are assessed. In general, a score of 0 means no manifestations/symptoms, a score of 1 indicates no significant impairment of function or activities of daily living (ADL), a score of 2 reflects significant impairment of ADL but no major disability, and a score of 3 indicates significant impairment of ADL with major disability. The scoring is conducted in the clinic and the only mandated laboratory tests for its completion are liver function tests, although pulmonary function tests are collected if available. An example of the 0-3 organ severity scoring is shown for skin (Figure 1). Clinicians provided the organ scoring information but global severity (mild, moderate, severe) was then calculated from these scores by computer algorithm according to the number and severity of organs reported. Mild disease was 1 or 2 organs (except lung) with score 1. Moderate disease was 3 or more organs with score 1 or lung score 1, or 1 or more organs with score 2. Severe disease was any organ with a score 3 or lung score 2. Lung dysfunction was treated differently than other organ dysfunction based on studies reporting higher mortality.⁷ In a single case, a patient was asymptomatic at enrollment (score 0 on all organs) although she was still on immunosuppression. This patient was combined with the mild global severity group for analysis.

Statistical considerations

For purposes of the main analyses, global severity scoring at the time of study enrollment (whether incident or prevalent case) was used (n = 298 adults). Global severity from either the enrollment visit or follow-up visits (n = 738) was used in the subsequent analyses. Statistics were descriptive for percentages and frequencies of categorical variables. For Karnofsky performance status (KPS), data were divided into tertiles for analysis.

The contribution of individual organ scores to the global severity score at each visit was quantified by calculating whether knowledge of the individual organ score was contributory (ie, helped determine the global

severity score) or necessary (ie, was the only score determining the global severity score).

Previously reported risk factors associated with the development of chronic GVHD as well as risk factors previously identified to be associated with increased nonrelapse mortality in patients with chronic GVHD were analyzed for their association with global severity of chronic GVHD using logistic regression. Generalized estimating equation (GEE) methods available for logistic regression in SAS Proc Genmod were used to adjust for baseline characteristics and variables that could vary over repeated observations per patient (n = 293 adults, 725 assessments). Missing data (n = 5 patients, n = 13 assessments) are attributable to missing key clinical information.

Nonrelapse mortality was defined as death without prior relapse. Survival was calculated from the time of enrollment, with patients censored at date last known alive. Cox regression was used for hazard ratio (HR) analysis of nonrelapse mortality and survival relative to severity and other risk factors.

Results

Patient and transplantation characteristics of the 298 participants included in this analysis are shown in Table 1. Chronic GVHD characteristics are summarized in Table 2.

At the time of study enrollment, global chronic GVHD severity according to NIH Consensus Criteria was calculated from reported data as mild in 10% (n = 32), moderate in 59% (n = 175), and severe in 31% (n = 91; Figure 1) Severity distribution was similar across incident (chronic GVHD diagnosis within 3 months of enrollment) and prevalent cases (P = .35). Skin, mouth, and liver were most commonly involved in mild chronic GVHD, while moderate and severe chronic GVHD often involved the skin, mouth, eye, liver, and lung (Table 2). Overall, the global severity assignments were attributable to lung (45%), skin (36%), eye (25%), mouth (15%), liver (12%), joint (11%), genital tract (6%), and GI tract (5%; column 2, Table 3). This means that the lung score contributed to the global severity score in 45% of the visits, even though the global severity could also have been determined from other organ involvement. In the analysis of necessary organ scoring, where the percentages represent the time that the organ must be scored to ascertain the correct global severity (column 3,

Table 1. Patient and transplantation characteristics at enrollment

	Severity at enrollment, N = 298								
	Mild, n = 32			Moderate, n = 175			Severe, n = 91		
	n (%)	n	Median (range)	n (%)	n	Median (range)	n (%)	n	Median (range)
Age at transplantation, y		32	51.5 (23.6-67.9)		175	52.0 (19.9-74.0)		91	51.8 (19.0-78.9)
Age at enrollment, y		32	53.0 (26.0-68.0)		175	53.0 (22.0-74.0)		91	53.0 (20.0-79.0)
Sites		32			175			91	
Fred Hutchinson Cancer Research Center	11 (35)			95 (54)			51 (56)		
University of Minnesota	6 (19)			22 (13)			7 (8)		
Dana-Faber Cancer institute	3 (9)			18 (10)			14 (15)		
Stanford University Medical Center	9 (28)			29 (17)			10 (11)		
Vanderbilt University Medical Center	3 (9)			11 (6)			9 (10)		
Donor sex		32			174			90	
Female into male	18 (56)			42 (24)			27 (30)		
Other	14 (44)			132 (76)			63 (70)		
White, non-Hispanic	30 (94)	32		153 (87)	175		82 (90)	91	
Diagnosis		32			175			91	
Acute leukemia (AML/ALL)	13 (40)			85 (49)			37 (41)		
Chronic leukemia (CML/CLL)	4 (13)			23 (13)			12 (13)		
MDS	4 (13)			30 (17)			15 (17)		
NHL/HD	10 (31)			22 (12)			21 (23)		
MM	1 (3)			11 (6)			5 (5)		
AA	0 (0)			1 (1)			0 (0)		
Other	0 (0)			3 (2)			1 (1)		
Disease stage		32			175			90	
Early	12 (38)			59 (34)			28 (31)		
Intermediate	18 (56)			75 (43)			42 (47)		
Advanced	2 (6)			41 (23)			20 (22)		
Graft source		32			175			91	
Peripheral blood	26 (82)			158 (90)			83 (91)		
BM	3 (9)			10 (6)			6 (7)		
Cord blood	3 (9)			7 (4)			2 (2)		
Transplant type		32			175			91	
Myeloablative	16 (50)			105 (60)			49 (54)		
Nonmyeloablative	16 (50)			70 (40)			42 (46)		
Donor type		32			173			91	
HLA-matched relative	14 (44)			88 (51)			38 (42)		
HLA-mismatched relative	0 (0)			4 (2)			2 (2)		
Unrelated donor	18 (56)			81 (47)			51 (56)		
CMV status (donor/patient)		32			174			90	
+/+	8 (25)			56 (32)			26 (29)		
+/-	4 (13)			13 (7)			14 (16)		
-/+	9 (28)			50 (29)			23 (25)		
-/-	11 (34)			55 (32)			27 (30)		
In vivo T-cell depletion (anti-thymocyte globulin)	5 (16)	32		8 (5)	175		5 (5)	91	
GVHD prophylaxis regimens		32			175			91	
CSA or tacrolimus + MTX ± others	16 (50)			82 (47)			44 (48)		
CSA or tacrolimus + MMF or sirolimus	7 (22)			74 (42)			37 (41)		
Tacrolimus + MMF + sirolimus	2 (6)			6 (3)			2 (2)		
Others	2 (6)			5 (3)			3 (3)		

AML indicates acute myeloid leukemia; ALL, acute lymphoblastic leukemia; CML, chronic myeloid leukemia; CLL, chronic lymphocytic leukemia; MDS, myelodysplastic syndrome; NHL, non-Hodgkin lymphoma; HD, Hodgkin lymphoma; MM, multiple myeloma; AA, aplastic anemia; CSA, cyclosporine A; MTX, methotrexate; and MMF, mycophenolate mofetil.

Table 3), the order was identical but frequencies were lower. The 5 least influential organs each accounted for < 6% of the global severity scores, but failure to score them would mean that global severity would be underestimated in 16% of visits.

The moderate global severity category was heterogeneous with 18 (10%) classified as moderate because of 3 or more score 1 organ manifestations and 108 (62%) classified as moderate because at least 1 organ had a score of 2 or lung score of 1. Forty-nine patients (28%) would have been classified as moderate by either criterion.

Patients assigned to the severe global category (n = 91) often had score 3 skin (n = 39, > 50% BSA, or deep sclerotic features,

or impaired mobility) or score 2-3 lung involvement (n = 44, FEV1 < 60% or lung function score 6 or greater or shortness of breath after walking on flat ground or at rest), accounting for 85% of assignments to the severe category. Scores of 3 in the mouth (n = 6), eye (n = 7), GI tract (n = 2), joints (n = 4), or genital tract (n = 4, females only) occurred in 3%-11% of patients in this category (Figure 1). There was no evidence that a specific pattern of organ involvement in the severe category was associated with nonrelapse mortality (P = .94) or survival (P = .85).

Global severity of chronic GVHD was not associated with previously reported risk factors for chronic GVHD onset such as

Table 2. Chronic GVHD characteristics at enrollment

	Severity at enrollment, N = 298								
	Mild (N = 32)			Moderate (N = 175)			Severe (N = 91)		
	n (%)	n	Median (range)	n (%)	n	Median (range)	n (%)	n	Median (range)
Months from transplantation to cGVHD diagnosis		32	7.4 (1.4-23.7)		175	7.2 (2.3-33.3)		91	7.0 (2.0-28.6)
Months from cGVHD diagnosis to enrollment		32	2.2 (0-31.3)		175	3.3 (0-35.3)		91	2.0 (0-26.9)
Comorbidities at enrollment		32	2 (0-7)		175	3 (0-9)		91	3 (0-10)
Type		32			175			91	
Incident	19 (59)			87 (50)			53 (58)		
Prevalent	13 (41)			88 (50)			38 (42)		
Chronic GVHD type		31			158			82	
Overlap acute and chronic GVHD	15 (48)			66 (42)			40 (49)		
Classic chronic GVHD	16 (52)			92 (58)			42 (51)		
Total number of organs involved		32	2 (0-2)		175	3 (1-7)		91	4 (1-7)
Organs involved									
Skin	13 (41)	32		105 (60)	175		70 (77)	91	
Mouth	14 (44)	32		113 (65)	175		57 (63)	91	
Eye	3 (9)	32		99 (57)	175		49 (54)	91	
Gastrointestinal	3 (9)	32		49 (28)	175		28 (31)	91	
Liver	13 (41)	32		83 (48)	173		53 (58)	91	
Joint	3 (9)	32		46 (26)	175		29 (32)	91	
Genital	0 (0)	26		23 (14)	163		14 (16)	89	
Lung	0 (0)	32		90 (51)	175		58 (64)	91	
Platelet count at onset, 10⁹/L		32			172			88	
< 100	7 (22)			39 (23)			24 (27)		
≥ 100	25 (78)			133 (77)			64 (73)		
Total serum bilirubin, mg/dL		32	0.5 (0.2-2.1)		171	0.6 (0.1-19.7)		88	0.7 (0.3-15.1)
Karnofsky performance status at enrollment		26			150			80	
≤ 70	9 (35)			45 (30)			38 (48)		
80	4 (15)			36 (24)			13 (16)		
90-100	13 (50)			69 (46)			29 (36)		
Unusual manifestations									
Pleural effusion(s)	0 (0)	32		3 (2)	175		3 (3)	91	
Bronchiolitis obliterans syndrome	0 (0)	32		8 (5)	175		9 (10)	91	
COP (formally bronchiolitis obliterans organizing pneumonia)	1 (3)	32		3 (2)	175		0 (0)	91	
Nephrotic syndrome	0 (0)	32		0 (0)	175		0 (0)	91	
Malabsorption	0 (0)	32		1 (1)	175		0 (0)	91	
Esophageal stricture or web	0 (0)	32		1 (1)	175		1 (1)	91	
Ascites (serositis)	0 (0)	32		1 (1)	175		1 (1)	91	
Myasthenia gravis	0 (0)	32		0 (0)	175		0 (0)	91	
Peripheral neuropathy	1 (3)	32		11 (6)	175		7 (8)	91	
Polymyositis	0 (0)	32		0 (0)	175		0 (0)	91	
Pericardial effusion	0 (0)	32		1 (1)	175		0 (0)	91	
Cardiomyopathy	0 (0)	32		0 (0)	175		0 (0)	91	
Cardiac conduction defects	0 (0)	32		1 (1)	175		2 (2)	91	
Coronary artery involvement	0 (0)	32		0 (0)	175		0 (0)	91	

cGVHD indicates chronic GVHD; and COP, cryptogenic organizing pneumonia.

Table 3. The importance of scoring individual organs to determine the global severity at each visit

Organ	Contributed to global severity score, %	Necessary to calculate global severity score, %
Lung	45	22
Skin	36	15
Eye	25	9
Mouth	15	6
Liver	12	5
Joint	11	2
Genital	6	2
Gastrointestinal	5	1

Organs "contributed" if they were used in the algorithm to calculate global severity but other organs also contributed. Organs were "necessary" to score if they solely determined the global severity score.

older age, female donors for male patients, unrelated or HLA-mismatched donors, conditioning intensity, peripheral blood grafts, prior CMV infection, underlying disease, disease status, or prior acute GVHD,⁸⁻¹⁰ nor with previously defined risk factors for mortality in patients with chronic GVHD, such as time to onset of chronic GVHD or thrombocytopenia (< 100 × 10⁹/L). Of the evaluated factors, only KPS ≤ 70 at time of chronic GVHD diagnosis was associated with higher global severity (Table 4).

The median follow-up of survivors was 18.5 months (range 2-41 months). Higher NIH global severity at enrollment was associated with higher nonrelapse mortality and lower overall survival, overall *P* values < .0001 (Figure 2). Thrombocytopenia (platelets < 100 × 10⁹/L) was also associated with nonrelapse mortality (HR 3.4: 1.7-6.7, *P* = .001) and survival (HR

Table 4. Risk factors for calculated NIH global severity

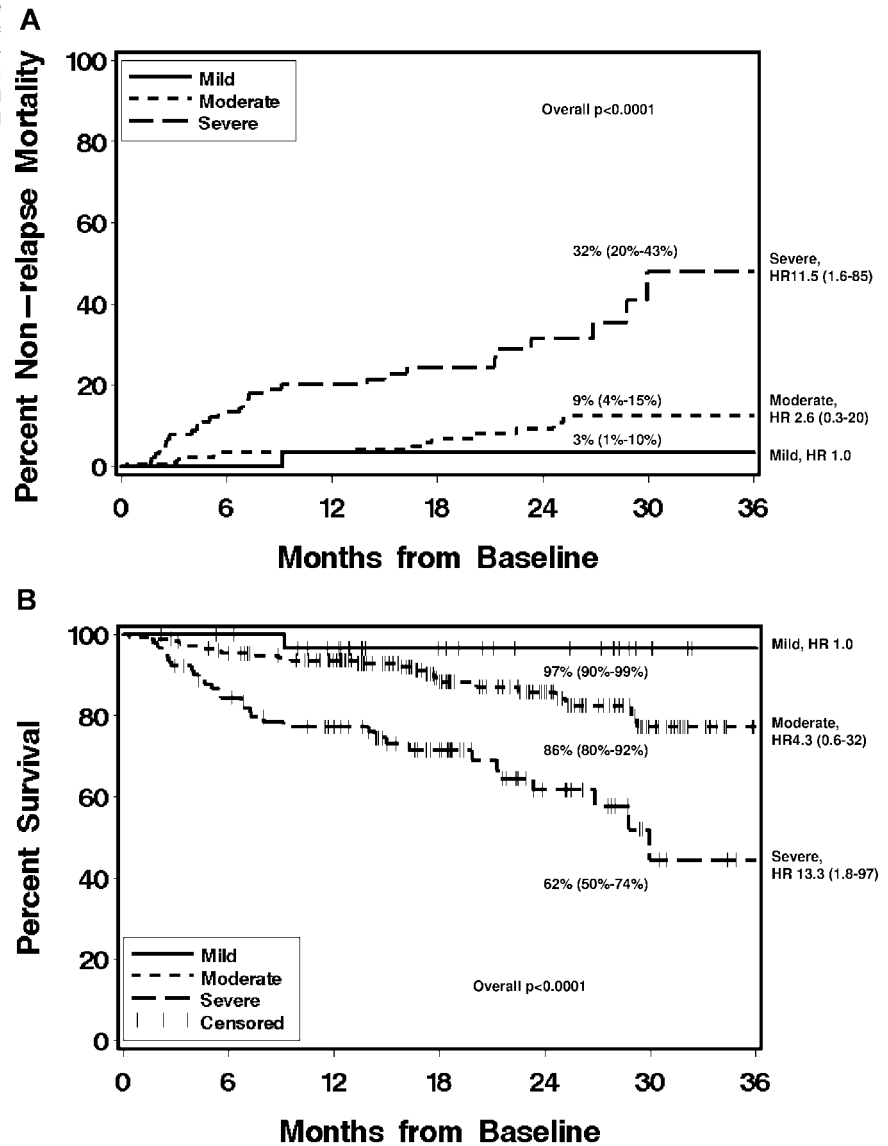
	Moderate/severe versus mild (625 vs 100 assessments)		Severe versus mild/moderate (195 vs 530 assessments)	
	OR (95% CI)	P	OR (95% CI)	P
Age at transplantation, y				
Adult < 50	1.0		1.0	
Adult ≥ 50	0.96 (0.5-1.9)	.91	0.96 (0.6-1.6)	.86
Sex				
Male	1.0		1.0	
Female	0.81 (0.4-1.7)	.58	0.95 (0.6-1.6)	.85
Donor/patient sex				
Other	1.0		1.0	
Female to male	0.46 (0.2-1.0)	.04	0.91 (0.5-1.6)	.75
Donor				
HLA-matched related	1.0		1.0	
HLA-matched URD	0.74 (0.3-1.6)	.43	1.10 (0.6-1.9)	.73
HLA-mismatched	0.66 (0.3-1.6)	.37	1.41 (0.6-3.1)	.40
Conditioning				
Myeloablative	1.0		1.0	
RIC/NMA	1.05 (0.5-2.2)	.89	1.36 (0.7-2.6)	.36
Stem cell source				
Peripheral blood	1.0		1.0	
BM	0.68 (0.2-2.1)	.55	1.38 (0.5-3.8)	.56
Cord blood	2.70 (0.7-11)	.17	0.84 (0.2-4.2)	.83
Patient CMV				
-	1.0		1.0	
+	0.94 (0.5-1.7)	.84	0.92 (0.5-1.6)	.76
Donor CMV				
-	1.0		1.0	
+	1.04 (0.6-1.9)	.90	1.01 (0.6-1.7)	.98
Diagnosis				
ALL/AML	1.0		1.0	
CLL/CML	0.40 (0.1-1.1)	.13	0.78 (0.3-1.9)	.56
MDS	0.89 (0.4-2.2)	.80	0.84 (0.4-1.8)	.65
HD/NHL	0.36 (0.2-0.8)	.03	1.16 (0.5-2.7)	.72
Other	0.45 (0.2-1.3)	.18	0.40 (0.2-1.0)	.05
Disease stage				
Early	1.0		1.0	
Intermediate	1.22 (0.6-2.6)	.60	1.12 (0.6-2.1)	.72
Advanced	1.47 (0.6-3.5)	.38	1.23 (0.6-2.4)	.55
KPS at chronic GVHD onset				
80+	1.0		1.0	
≤ 70	2.71 (1.4-5.1)	.001	2.19 (1.4-3.4)	.0007
Missing	0.95 (0.5-1.9)	.88	1.38 (0.8-2.3)	.24
Prior acute GVHD				
No	1.0		1.0	
Yes	1.55 (0.9-2.8)	.15	0.89 (0.5-1.5)	.67
Time to onset, mo				
< 6	1.0		1.0	
6-12	0.71 (0.4-1.2)	.23	0.81 (0.5-1.4)	.42
≥ 12	1.95 (0.6-6.0)	.20	1.08 (0.5-2.2)	.83
Platelets at onset, 10⁹/L				
≥ 100	1.0		1.0	
< 100	1.43 (0.6-3.6)	.42	1.50 (0.8-2.9)	.24
Time since onset				
Per month	1.00 (0.98-1.03)	.79	1.00 (0.98-1.03)	.66

NIH indicates National Institutes of Health; OR, odds ratio; CI, confidence interval; RIC/NMA, reduced-intensity conditioning/nonmyeloablative; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; CML, chronic myeloid leukemia; CLL, chronic lymphocytic leukemia; MDS, myelodysplastic syndrome; NHL, non-Hodgkin lymphoma; URD, unrelated donor; and HD, Hodgkin lymphoma.

3.1: 1.7-5.6, $P = .0006$). KPS ≤ 70 at time of chronic GVHD diagnosis was associated with survival (HR 2.1: 1.2-3.8, $P = .05$). Nonrelapse mortality and survival were not associated with donor type, recipient age, or disease status (Table 5) or with incident or prevalent status or time from transplantation. Two-year nonrelapse mortality was 3% (95% CI, 1%-10%), 9%

(4%-15%), and 32% (20%-43%), and 2-year survival was 97% (95% CI, 90%-99%), 86% (80%-92%), and 62% (50%-74%) for mild, moderate, and severe global severity, respectively. The median survival for patients with severe chronic GVHD according to NIH consensus criteria was 30 months, while it has not been reached for patients with mild or moderate chronic GVHD.

Figure 2. Please provide a brief inclusive title to the Figure 2 legend that does not reference specific panel labels. Cumulative incidence of nonrelapse mortality (A) and Kaplan-Meier plot of overall survival (B) according to NIH global severity at enrollment, and showing 2-year estimates (95% CI). HR, hazard ratio and 95% CI, compared with mild chronic GVHD.



Discussion

We analyzed the spectrum of organ involvement and global and organ-specific chronic GVHD severity in 298 adult patients enrolled in the Chronic GVHD Consortium and report several findings. First, the distribution of NIH global severity scores was higher than we expected and did not differ according to duration of chronic GVHD. Second, the moderate chronic GVHD category, which comprised over half ($n = 175, 59\%$) of the patients, appears to be quite heterogeneous and is defined primarily by patients who have at least 1 organ of moderate severity or lung score of 1. Additional refinements to the global severity scoring may be able to distinguish prognostically different subgroups within the moderate category.

Higher NIH global severity was attributable both to a greater number of organs involved and more severe individual organ scoring, with lung, skin, or eye involvement contributing to the global severity score in the majority of visits. Although we had sought to identify at least one organ system that could be deleted from the scoring system without compromising the global severity

calculation, we were not able to do so. The 5 least common organs still should be scored because otherwise global severity would be underestimated in 16% of visits.

Our multivariate analysis revealed no association of NIH global severity categories and previously defined risk factors for development of chronic GVHD or nonrelapse mortality,^{1,3,11-16} other than a low KPS at diagnosis of chronic GVHD. These results suggest that most factors which predict onset and prognosis of chronic GVHD may be different from those that determine functional impairment and symptoms in individual organs. Our observation that $KPS \leq 70\%$ at chronic GVHD diagnosis was associated with moderate-severe chronic GVHD concurrently or at any subsequent assessment point may just reflect functional impairment, such as shortness of breath with exertion, being incorporated into the organ scoring definitions. Low KPS has been previously identified as a risk factor for mortality in patients with chronic GVHD.^{1,3} There is not yet enough data in the current cohort to know whether certain organ involvements within the global severity categories are most predictive of survival because we could not detect any organ-specific associations in the severe category, but we might have had limited power. For example, eye and skin manifestations are

Table 5. Multivariate analysis of overall mortality and nonrelapse mortality

	Overall mortality (54 events)		Nonrelapse mortality (41 events)	
	HR (95% CI)	P	HR (95% CI)	P
NIH severity				
Mild (n = 32)	1.0		1.0	
Moderate (n = 175)	5.0 (0.7-38)		2.9 (0.4-22)	
Severe (n = 91)	13.1 (1.8-97)	< .0001	10.9 (1.5-81)	< .0001
Platelets, 10⁹/L				
≥ 100	1.0		1.0	
< 100	3.1 (1.7-5.6)	.0006	3.4 (1.7-6.7)	.001
KPS at onset				
80+	1.0		1.0	
≤ 70	2.1 (1.2-3.8)		2.0 (1.0-4.0)	
Missing	1.2 (0.5-3.1)	.05	1.5 (0.5-4.1)	.13
Donor				
Matched related	1.0		1.0	
Matched URD	0.8 (0.4-1.5)		0.7 (0.4-1.5)	
Mismatched	1.0 (0.5-2.0)	.80	0.7 (0.3-1.8)	.62
Age at transplantation, y				
Adult < 50	1.0		1.0	
Adult ≥ 50	1.2 (0.7-2.1)	.60	1.1 (0.6-2.1)	.83
Disease stage				
Early	1.0		1.0	
Intermediate	1.4 (0.7-2.7)		1.2 (0.6-2.7)	
Advanced	1.3 (0.6-2.9)	.63	1.7 (0.7-4.0)	.50

HR indicates hazard ratio; CI, confidence interval; NIH, National Institutes of Health; KPS, Karnofsky performance status; and URD, unrelated donor.

common and often severe but may not be associated with life-threatening complications in the same direct way as lung dysfunction.

Retrospective series have reported mixed results in assessing whether NIH chronic GVHD global severity is associated with overall survival, where no correlation was found in one series,¹⁷ while worse survival was reported in other series.¹⁸⁻²¹ These studies were all conducted from chart review rather than prospectively collected information. Our results are based on prospectively collected data and show that global chronic GVHD severity scores calculated according to the NIH suggested scoring algorithm do have prognostic significance. Severe global chronic GVHD was associated with higher nonrelapse mortality and lower survival, with a median survival of 30 months. There was a trend for moderate chronic GVHD to have a higher nonrelapse mortality and lower survival than patients with mild chronic GVHD but this is not statistically significant yet, perhaps because of limited sample size, population heterogeneity, and relatively few fatal events during the current period of follow-up. We do not yet have enough cases of recurrent malignancy to analyze whether NIH chronic GVHD severity is associated with the graft-versus-tumor effect.²²

Our study has several limitations. First, while we are reporting a very large cohort of patients, these are all adults and reflect the transplantation practices and clinical assessments at a limited number of large institutions with a specific interest in chronic GVHD. Notably, children, ethnic and racial minorities, and other important subgroups are absent or underrepresented in this cohort. Second, because an indication for systemic treatment was required for enrollment into the cohort, patients with very mild chronic GVHD who only needed topical therapy are not represented. Third, because of the extensive nature of the required NIH data elements, we were not able to collect additional clinical information that could have been used to improve the organ-specific grading. The provider assessment is already long, requires specific training, and is challenging in the context of a busy clinical practice. Fourth, our median follow-up of survivors (18.5 months) is still short. Al-

though we were able to demonstrate statistical differences in nonrelapse mortality and overall survival, additional follow-up will allow more nuanced analyses and possible refinements to the global scoring system. Lastly, although individual institutions have collected some biologic samples on study participants, there is no standardized biologic repository across the consortium. Biomarker discovery in chronic GVHD populations which are well characterized phenotypically according to NIH criteria may help elucidate the biologic pathogenesis of the subtypes of chronic GVHD, potentially allowing development of targeted therapies.

In conclusion, we recommend that all organs continue to be scored according to NIH criteria in studies focused on chronic GVHD incidence and severity although lung, skin, and eye scores contribute the most to global severity scoring. The current NIH global scoring system has prognostic significance and appears to accurately reflect different risks of nonrelapse mortality and overall survival. No apparent clinical factors predict patients who will have severe chronic GVHD, but this group, which comprised 31% of our cohort, had a median survival of only 30 months justifying study of aggressive new interventions to improve survival in this very high-risk group.

Acknowledgments

This work was supported by National Institutes of Health/National Cancer Institute grant CA 118953.

Authorship

Contribution: S.A., M.J., C.C., M.A., D.J.W., M.E.D.F., P.J.M., and S.J.L. contributed clinical data; B.S. and X.C. performed statistical analysis; S.A., M.J., and S.J.L. designed research and drafted the

manuscript; and all authors contributed to analysis and interpretation of data and critical review of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Stephanie J. Lee, MD, MPH, Clinical Research Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, D5-290, Seattle, WA 98109; e-mail: sjlee@fhcrc.org.

References

- Lee SJ, Klein JP, Barrett AJ, et al. Severity of chronic graft-versus-host disease: association with treatment-related mortality and relapse. *Blood*. 2002;100(2):406-414.
- Pidala J, Kurland B, Chai X, et al. Patient reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. *Blood*. 2011;117(17):4651-4657.
- 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(2):204-217.
- Filipovich AH, Weisdorf D, Pavletic S, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. *Biol Blood Marrow Transplant*. 2005;11(12):945-956.
- Schultz KR, Miklos DB, Fowler D, et al. Toward biomarkers for chronic graft-versus-host disease: National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: III. Biomarker Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(2):126-137.
- Pavletic SZ, Martin P, Lee SJ, et al. Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: IV. Response Criteria Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(3):252-266.
- Walter EC, Orozco-Levi M, Ramirez-Sarmiento A, et al. Lung function and long-term complications after allogeneic hematopoietic cell transplant. *Biol Blood Marrow Transplant*. 2010;16(1):53-61.
- Atkinson K, Horowitz MM, Gale RP, et al. Risk factors for chronic graft-versus-host disease after HLA-identical sibling bone marrow transplantation. *Blood*. 1990;75(12):2459-2464.
- Ringden O, Paulin T, Lonnqvist B, Nilsson B. An analysis of factors predisposing to chronic graft-versus-host disease. *Exp Hematol*. 1985;13(10):1062-1067.
- Storb R, Prentice RL, Sullivan KM, et al. Predictive factors in chronic graft-versus-host disease in patients with aplastic anemia treated by marrow transplantation from HLA-identical siblings. *Ann Intern Med*. 1983;98(4):461-466.
- Akpek G, Zahurak ML, Piantadosi S, et al. Development of a prognostic model for grading chronic graft-versus-host disease. *Blood*. 2001;97(5):1219-1226.
- Arora M, Burns LJ, Davies SM, et al. Chronic graft-versus-host disease: a prospective cohort study. *Biol Blood Marrow Transplant*. 2003;9(1):38-45.
- Flowers ME, Inamoto Y, Carpenter PA, et al. Comparative analysis of risk factors for acute graft-versus-host disease and for chronic graft-versus-host disease according to National Institutes of Health consensus criteria. *Blood*. 2011;117(11):3214-3219.
- Stewart BL, Storer B, Storek J, et al. Duration of immunosuppressive treatment for chronic graft-versus-host disease. *Blood*. 2004;104(12):3501-3506.
- Vigorito AC, Campregher PV, Storer BE, et al. Evaluation of NIH consensus criteria for classification of late acute and chronic GVHD. *Blood*. 2009;114(3):702-708.
- Wingard JR, Piantadosi S, Vogelsang GB, et al. Predictors of death from chronic graft-versus-host disease after bone marrow transplantation. *Blood*. 1989;74(4):1428-1435.
- Jagasia M, Giglia J, Chinratanalab W, et al. Incidence and outcome of chronic graft-versus-host disease using National Institutes of Health consensus criteria. *Biol Blood Marrow Transplant*. 2007;13(10):1207-1215.
- Arora M, Nagaraj S, Witte J, et al. New classification of chronic GVHD: added clarity from the consensus diagnoses. *Bone Marrow Transplant*. 2009;43(2):149-153.
- Cho BS, Min CK, Eom KS, et al. Feasibility of NIH consensus criteria for chronic graft-versus-host disease. *Leukemia*. 2009;23(1):78-84.
- Kim DY, Lee JH, Kim SH, et al. Reevaluation of the National Institutes of Health criteria for classification and scoring of chronic GVHD. *Bone Marrow Transplant*. 2010;45(7):1174-1180.
- Perez-Simon JA, Encinas C, Silva F, et al. Prognostic factors of chronic graft-versus-host disease following allogeneic peripheral blood stem cell transplantation: the national institutes health scale plus the type of onset can predict survival rates and the duration of immunosuppressive therapy. *Biol Blood Marrow Transplant*. 2008;14(10):1163-1171.
- Thepot S, Zhou J, Perrot A, et al. The graft-versus-leukemia effect is mainly restricted to NIH-defined chronic graft-versus-host disease after reduced intensity conditioning before allogeneic stem cell transplantation. *Leukemia*. 2010;24(11):1852-1858.