

# Severity of chronic graft-versus-host disease: association with treatment-related mortality and relapse

Stephanie J. Lee, John P. Klein, A. John Barrett, Olle Ringden, Joseph H. Antin, Jean-Yves Cahn, Matthew H. Carabasi, Robert Peter Gale, Sergio Giralt, Gregory A. Hale, Osman Ilhan, Philip L. McCarthy, Gerard Socie, Leo F. Verdonck, Daniel J. Weisdorf, and Mary M. Horowitz

**Chronic graft-versus-host disease (cGVHD) is the leading cause of late treatment-related deaths among recipients of allogeneic bone marrow and blood transplants. However, cGVHD is also associated with fewer relapses. We sought to determine whether severity of cGVHD predicts the magnitude of these effects. One impediment to such an analysis is the current limited/extensive grading system for cGVHD because this classification was designed to identify patients likely to benefit from systemic immune suppression and does not capture the severity of multiorgan involvement. We, therefore, first developed a grading system predictive**

**for survival by using data from 1827 HLA-matched sibling allotransplant recipients reported to the International Bone Marrow Transplant Registry (IBMTR). We found Karnofsky performance score, diarrhea, weight loss, and cutaneous and oral involvement to be independent prognostic variables, from which we generated a grading scheme. We tested this scheme, the limited/extensive classification system, and a classification based on clinical impression of overall cGVHD severity (mild/moderate/severe) in parallel analyses of 1092 HLA-matched sibling transplant recipients from the IBMTR and 553 recipients of unrelated donor marrow from**

**the National Marrow Donor Program. Presence of cGVHD was associated with fewer relapses (relative risk [RR], 0.5-0.6) but more treatment-related mortality (RR, 1.8-2.8) in the 3 analyses. No grading scheme correlated cGVHD severity with relapse rates, but all schemes predicted treatment-related mortality. Survival and disease-free survival of the most favorable cGVHD group in each scheme were similar, or better, than those of patients without cGVHD; these patients may not need aggressive or extended immune suppression. (Blood. 2002;100:406-414)**

© 2002 by The American Society of Hematology

## Introduction

Chronic graft-versus-host disease (cGVHD) is an important complication of allogeneic stem cell transplantation. From 30% to 70% of allograft recipients develop cGVHD<sup>1,2</sup> that is associated with decreased quality of life,<sup>3</sup> impaired functional status,<sup>4,5</sup> need for extended immune suppression, and impaired survival.<sup>6-10</sup> Several trends in allogeneic transplantation, including use of transplants in older patients, use of peripheral blood cell grafts,<sup>11-17</sup> and use of unrelated or HLA-mismatched donors, may increase incidence of cGVHD.

Despite its adverse effects, cGVHD is associated with fewer leukemia relapses. This effect is thought to reflect a graft-versus-leukemia effect comparable or greater than that ascribed to acute GVHD.<sup>18-23</sup> Consequently, the influence of cGVHD on survival reflects the balance of its negative (increased treatment-related mortality) and positive (fewer relapses) effects. Describing this relationship requires an accurate measure of cGVHD severity.

The current grading scheme for cGVHD severity was proposed in 1980 based on data on 20 subjects. Chronic GVHD was classified as

From the Graft vs Host Disease Working Committee of the International Bone Marrow Transplant Registry, Health Policy Institute, Medical College of Wisconsin, Milwaukee; the Coordinating Center of the National Marrow Donor Program, Minneapolis, MN; Dana-Farber Cancer Center, Boston, MA; National Heart, Lung, and Blood Institute, Bethesda, MD; Huddinge Hospital, Huddinge, Sweden; Hopital Jean Minjot, Besancon, France; University of Southern California/Norris Cancer Center, Los Angeles; Center for Advanced Studies in Leukemia, Los Angeles, CA; MD Anderson Cancer Center, Houston, TX; St Jude Children's Research Hospital, Memphis, TN; Ilni Sina Hospital, Ankara, Turkey; Roswell Park Cancer Institute, Buffalo, NY; Hôpital Saint Louis, Paris, France; University Hospital Utrecht, The Netherlands; and Fairview University Medical Center, Minneapolis, MN.

Submitted July 13, 2001; accepted March 7, 2002.

Supported in part by National Institutes of Health Grant no. CA75267-03 and the Amy Strelzer-Manasevit Scholars Program. National Marrow Donor Program, the Health Resources and Services Administration no. 240-97-0036. Additional support provided by Public Health Service Grants P01-CA-40053 and U24-76518 from the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, and the National Heart, Lung, and Blood Institute of the U.S. Department of Health and Human Services; and grants from Alpha Therapeutic Corporation; Amgen, Inc; Anonymous; Baxter Fenwal; Berlex Laboratories; BioWhittaker, Inc; Blue Cross and Blue Shield Association; Lynde and Harry Bradley Foundation; Bristol-Myers Squibb

Company; Cell Therapeutics, Inc; Centeon; Center for Advanced Studies in Leukemia; Chimeric Therapies; Chiron Therapeutics; Charles E. Culpeper Foundation; Eleanor Naylor Dana Charitable Trust; Eppley Foundation for Research; Fromstein Foundation; Genentech, Inc; Human Genome Sciences; Immunex Corporation; Kettering Family Foundation; Kirin Brewery Company; Robert J. Kleberg Jr and Helen C. Kleberg Foundation; Herbert H. Kohl Charities, Inc; Nada and Herbert P. Mahler Charities; Milstein Family Foundation; Milwaukee Foundation/Elsa Schoeneich Research Fund; NeXstar Pharmaceuticals, Inc; Samuel Roberts Noble Foundation; Novartis Pharmaceuticals; Orphan Medical; Ortho Biotech, Inc; John Oster Family Foundation; Jane and Lloyd Pettit Foundation; Alirio Piffier Bone Marrow Transplant Support Association; Pfizer, Inc; RGK Foundation; Rockwell Automation Allen Bradley Company; Roche Laboratories; SangStat Medical Corporation; Schering AG; Schering-Plough Oncology; Searle; SEQUUS Pharmaceuticals; SmithKline Beecham Pharmaceutical; Stackner Family Foundation; Starr Foundation; Joan and Jack Stein Foundation; SyStemix; United Resource Networks; and Wyeth-Ayerst Laboratories.

**Reprints:** Stephanie Lee, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; e-mail: stephanie\_lee@dfci.harvard.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.

© 2002 by The American Society of Hematology

limited (involving only localized skin and/or liver) or extensive (generalized skin or limited disease plus involvement of other organs) involvement. This system was developed primarily to distinguish patients requiring systemic immune suppression from those for whom local care might suffice. The heterogeneity of organ involvement, clinical severity, and prognosis, especially within the extensive category, was acknowledged but purposefully not incorporated into the grading scale.<sup>24</sup> Both the International Bone Marrow Transplant Registry (IBMTR) and the National Marrow Donor Program (NMDP) use this grading system to report cGVHD severity.

There are several grading schemes that predict survival of patients with cGVHD.<sup>6,8,25,26</sup> In these grading schemes, poor prognostic variables include lichenoid skin changes, extensive skin involvement (> 50% body surface area), elevated bilirubin, progressive onset, thrombocytopenia, and prior steroid refractory/dependent acute GVHD. We were unable to calculate severity scores with the use of these grading schemes in our study because several variables (lichenoid skin changes, extent of skin involvement, prior steroid refractory/dependent acute GVHD) are unavailable in the IBMTR and NMDP databases.

We analyzed data from the IBMTR and the NMDP to develop and, hopefully, validate a new cGVHD severity scoring system based on patterns of organ involvement. We then used this system and the previous limited/extensive and clinical grading of cGVHD (mild/moderate/severe) to study the correlation between cGVHD severity and treatment-related mortality, relapse, disease-free survival, and survival.

## Patients, materials, and methods

### Patients

Three independent data sets were used (IBMTR-1, IBMTR-2, and NMDP). Each group represented nonoverlapping populations with data obtained by using different versions of data collection forms. IBMTR-1 served as the training set for the new cGVHD severity score, whereas IBMTR-2 and NMDP were used to test validity. All subjects met the following eligibility criteria: (1) acute myelogenous leukemia, acute lymphoblastic leukemia, or chronic myelogenous leukemia (CML); (2) non-T-cell-depleted transplant from an HLA-matched sibling or unrelated donor (matched at HLA-A, -B,

**Table 1. Characteristics of the training and testing populations**

Variable	IBMTR-1 N = 1827	IBMTR-2 N = 1092	NMDP N = 553
Designation	Training	Validation	Validation
Version of data collection form	089	095	NMDP
Donor	HLA-matched siblings	HLA-matched siblings	HLA-matched unrelated donors
Median age, y (range)	34 (16-60)	34 (16-69)	34 (16-55)
Male, n (%)	1077 (59)	635 (58)	351 (63)
Disease, n (%)			
CML	908 (50)	532 (49)	399 (72)†
AML	634 (35)	381 (35)	96 (17)
ALL	285 (16)	179 (16)	58 (10)
Disease stage, n (%)			
Early	1405 (77)	844 (77)	318 (58)†
Intermediate	241 (13)	144 (13)	147 (27)
Advanced	181 (10)	104 (10)	84 (15)
Sex matching, n (%)			
M into M	625 (34)	350 (32)	222 (40)
M into F	416 (23)	237 (22)	106 (19)
F into M	452 (25)	285 (26)	129 (23)
F into F	334 (18)	220 (20)	96 (17)
CMV-positive patient or donor, n (%)	1419 (78)	825 (76)	361 (66)†
Total body irradiation, n (%)	963 (53)	462 (42)†	490 (89)†
Year of transplant, n (%)			
1990	453 (25)	34 (3)†	49 (9)†
1991	374 (21)	44 (4)	79 (14)
1992	382 (21)	86 (8)	120 (22)
1993	372 (20)	94 (9)	142 (26)
1994	246 (13)	151 (14)	163 (29)
1995	0 (0)	283 (26)	0 (0)
1996	0 (0)	253 (23)	0 (0)
1997	0 (0)	147 (13)	0 (0)
Acute GVHD, n (%)			
None	828 (45)	529 (49)	93 (17)*
I	487 (27)	250 (23)	103 (19)
II	354 (19)	202 (19)	126 (23)
III	120 (7)	68 (6)	166 (31)
IV	38 (2)	30 (3)	50 (9)
Missing, n		13	15
Median follow-up of survivors, d (range)‡	504 (101-2039)	455 (101-2926)	2200 (273-3354)
Cumulative incidence of cGVHD at 1 y, %	42	43	63†
Cumulative incidence of cGVHD at 2 y, %	45	45	67†

CML indicates chronic myelogenous leukemia; AML, acute myelogenous leukemia; ALL, acute lymphoblastic leukemia; and CMV, cytomegalovirus.

Note: statistical symbols for categorical variables are displayed next to the first line only.

\* $P < .001$  compared with training set.

† $P < .0001$  compared with training set.

‡Statistical testing not performed.

-DR by serologic or molecular testing); (3) age of 16 years or older; (4) transplantation after 1990; (5) acute GVHD prophylaxis with cyclosporine and methotrexate; and (6) disease-free survival 100 days or more.

### Data collection instrument and methods

Subject-, disease-, and transplant-related variables and outcomes were collected on standardized forms of the IBMTR and NMDP. The IBMTR is a voluntary working group of more than 350 transplant teams worldwide that contribute detailed data on their allogeneic transplants to the Statistical Center at the Medical College of Wisconsin. Participants are required to register all consecutive transplants. The IBMTR database includes information on 40% to 45% of all allogeneic transplant recipients since 1970. All patients are followed longitudinally for survival and relapse. However, the IBMTR uses 2 follow-up forms, the abbreviated and the comprehensive versions; only the comprehensive form collects cGVHD data. To ensure that late onset cGVHD was not missed and to maintain comparability of follow-up between patients who did and did not develop cGVHD, all patients were censored at the time of last comprehensive follow-up form completion even if more recent survival and relapse data were available from abbreviated forms. Computerized error checks, physician review of submitted data, and on-site audits of participating centers ensure data quality.

The NMDP, established in 1986, maintains a registry of more than 4 million volunteer stem cell donors, facilitates unrelated donor stem cell transplants, and performs research in the area of unrelated donor stem cell transplantation. Through its network of participating centers, the NMDP has facilitated more than 12 000 transplantations. The 128 participating transplant centers prospectively report recipient baseline and follow-up data

to the NMDP Coordinating Center in Minneapolis. Baseline and follow-up data on almost 85% of the NMDP-facilitated transplants have been reported to the NMDP database. Data quality is maintained through on-line computer validation and on-site data audits.

Relapse was defined by hematologic criteria for all diseases. Molecular and cytogenetic relapses for chronic myelogenous leukemia were not included because they are not 100% predictive for subsequent relapse and because frequency of monitoring likely varies by center. Chronic GVHD information was collected on 3 different versions of one data form and varied in the degree of detail requested. This variation necessitated combining or substituting some variables to apply the new severity score to the validation sets. These conversions are outlined in "Appendix" and were established before analysis. Clinical assessment of cGVHD severity (mild, moderate, or severe) was collected on IBMTR patients. However, no guidelines were provided to standardize this judgment.

The NMDP collects information at 100 days, 6 months, 1 year, and then annually after transplantation. Median time of completion of the first 3 follow-up forms is 100 days, 6 months, and 1 year, respectively. In contrast, the first IBMTR form is completed a median of 7.5 months after transplantation, and the second is completed approximately 19 months after transplantation. Any manifestations reported on the first 3 NMDP forms (comprising follow-up to 1 year) were considered "initial" manifestations in the calculation of the new severity score to be comparable with the IBMTR training set in which the score was developed.

Missing data are due to absent questions on the study form or incomplete reporting by centers. For example, data on bilirubin was not collected for the IBMTR-1 group and weight loss was not collected for the

**Table 2. Chronic graft-versus-host disease in the training and testing sets**

Variable	IBMTR-1 n = 727	IBMTR-2 n = 421	NMDP n = 378
Designation	Training	Validation	Validation
Median time to onset cGVHD, d (range)	136 (60-1327)	140 (60-1266)	122 (66-1127)
Type of onset, n (%)			
Progressive	157 (22)	132 (31)‡	73 (19)§
Interrupted	314 (43)	138 (33)	260 (69)
De novo	256 (35)	151 (36)	45 (12)
Prior acute GVHD, n (%)	470 (65)	270 (64)	333 (88)§
Median platelet count at onset × 10 <sup>9</sup> /L (range)	131 (5-999)	114 (0-548)‡	91 (1-448)§ substantial missing data
Karnofsky performance status at onset, median (range)	80 (20-100)	80 (20-100)	70 (20-100)§ substantial missing data
Total bilirubin > 2 mg/dL at onset§, n (%)	Not collected	55 (16) substantial missing data	21 (22) substantial missing data
Chronic GVHD, n (%)			
Limited	350 (48)	225 (54)	58 (15)§
Extensive	377 (52)	191 (46)	319 (85)
Severity at onset, n (%)			
Mild	381 (52)	215 (51)	Not collected
Moderate	247 (34)	142 (34)	
Severe	99 (14)	62 (15)	
New cGVHD grade, n (%)			
Low risk	365 (50)	256 (61)*	64 (34)†
Intermediate risk	232 (32)	98 (23)	79 (42)
High risk	130 (18)	67 (16)	45 (24)
Not able to classify, n			190
Treatment, %			
Cyclosporine	60	81§	93§
Tacrolimus	Not collected	4	Not collected
Topical steroids	15	18	17
Systemic steroids	81	84	95§
ATG/ALS	2	6†	6‡
Azathioprine	14	11	14
Thalidomide	9	8	12
Resolution of cGVHD  , n (%)	245 (34)	Not collected	Not collected
Time to resolution of cGVHD from onset date, d (range)	369 (79-1545)	Not collected	Not collected

ATG/ALS indicates antithymocyte globulin/antilymphocyte serum.

\**P* ≤ .05 compared with training set.

†*P* ≤ .01 compared with training set.

‡*P* ≤ .001 compared with training set.

§*P* ≤ .0001 compared with training set.

||Statistical testing not performed.

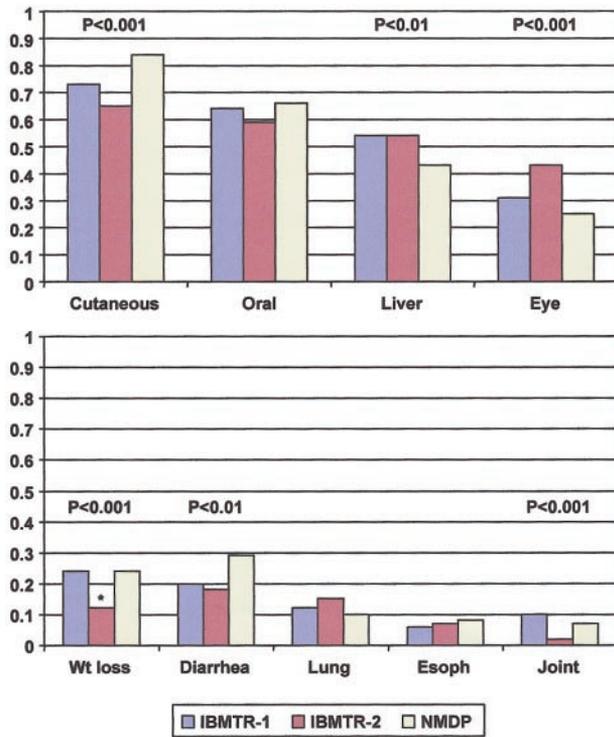


Figure 1. Spectrum of organ involvement reported in the IBMTR-1, IBMTR-2, and NMDP cohorts. \*Weight loss not collected in IBMTR-2, proportion represents patients with nausea or malabsorption.

IBMTR-2 cohort. The NMDP did not request Karnofsky performance score (KPS) before 1993 or collect a clinical impression of severity of cGVHD (mild/moderate/severe).

**Biostatistical analysis**

Descriptive statistics are reported for the 3 cohorts. The IBMTR-1 (training) group was used to develop the severity score that was tested in the IBMTR-2 and NMDP cohorts (validation sets).

Classifications of limited or extensive cGVHD were compared with reported organ involvement to assess whether this grading system was interpreted correctly by transplant centers. We assumed that involvement of only skin and/or liver was consistent with limited disease. We furthermore considered chronic diarrhea or eye, mouth, esophageal, joint, lung, genitourinary, or muscular involvement consistent with extensive disease whether or not there was skin and/or liver involvement. This algorithm misclassifies some patients with extensive cGVHD as having limited disease because patients with generalized skin involvement or liver cirrhosis actually have extensive disease.

**Development and validation of the new severity score**

Two proportional hazards models were created by using the training set: the first used a delayed onset model to predict overall survival from time of transplantation so that subjects entered the risk set when they developed cGVHD. A second model predicted survival time from date of cGVHD onset. Candidate variables included (1) type of onset (progressive, interrupted, de novo); (2) platelet count at onset ( $< 100 \times 10^9/L$  versus  $\geq 100 \times 10^9/L$ ); (3) KPS ( $< 80\%$  versus  $\geq 80\%$ ); (4) organ-specific involvement: skin, eye, mouth, esophagus, liver, joint, and lung; and (4) symptoms: diarrhea and weight loss. (Note: Bilirubin level was not collected on IBMTR-1 patients and, therefore, was not a potential predictor.) Time-varying covariates were used to represent organ involvement and symptoms. Of the baseline covariates (year of transplantation, disease type, disease stage [early, intermediate, advanced], donor-recipient sex-matching, age, cytomegalovirus serologic status, use of total body irradiation for conditioning, and whether or not acute GVHD occurred), only disease stage was found to predict survival in patients with cGVHD

and was included in all models. Patients were censored at time of second transplantation, donor lymphocyte infusion, or last follow-up if alive.

Proportional hazards models were fitted for different combinations of organ involvement, and those combinations present in at least 5 subjects retained for analysis. Combinations with overall survival curves statistically different than baseline ( $P < .05$ ) were identified and grouped.

Comparisons were made between the predictive ability of the new severity score and the limited/extensive and mild/moderate/severe grading scales as reported by the transplant centers using Akaike information criterion (AIC).<sup>27,28</sup> When models are run on identical subject populations, a lower AIC indicates a better fit and higher predictive ability.

**Comparison of cGVHD severity groups with patients who did not develop cGVHD**

Severity groups were compared with each other and with subjects without cGVHD by using time-varying covariates for the cGVHD severity score (patients without cGVHD formed the baseline group). Analyses were adjusted for disease type and disease stage, age, donor-recipient sex-matching, and whether or not prior acute GVHD occurred since these baseline covariates were associated with survival in the entire population. Possible interactions between cGVHD severity and the baseline covariates were evaluated by using interaction terms.

Although the cGVHD severity score was developed by using a survival end point, its predictive values for leukemia-free survival, treatment-related mortality, and relapse were also tested. Survival curves were presented by using left-truncated Kaplan-Meier plots. Relapse and treatment-related mortality were presented as cumulative incidences.<sup>29</sup>

**Results**

**Characteristics of the training and validation sets**

Table 1 shows the characteristics of the cohorts studied. Baseline subject- and transplant-related variables were similar among the 3 data sets, although there were some differences in disease distribution, proportion of recipient/donor cytomegalovirus-negative subjects, use of total body irradiation, and year of transplantation. The distribution of grades of acute GVHD was skewed upward in unrelated transplant recipients compared with sibling donor transplant recipients.

Because the IBMTR does not collect cGVHD data on its abbreviated follow-up forms, surviving recipients of sibling transplants were censored at date of last comprehensive reporting to ensure that cases of late onset cGVHD were not missed. Consequently, follow-up of IBMTR survivors was shorter (median, 1.4 and 1.2 years for the 2 groups) than NMDP survivors (median, 6.0 years,  $P < .0001$ ). We were not able to classify 190 (50%) of NMDP patients by the new score because forms were completed before a question on KPS data was added. At last follow-up there were relapses in 243 (13%) IBMTR-1 patients, 132 (12%) IBMTR-2 patients, and 64 (12%) NMDP patients. There was treatment-related mortality in 221 (12%) IBMTR-1 patients, 130 (12%) IBMTR-2 patients, and 256 (46%) NMDP patients.

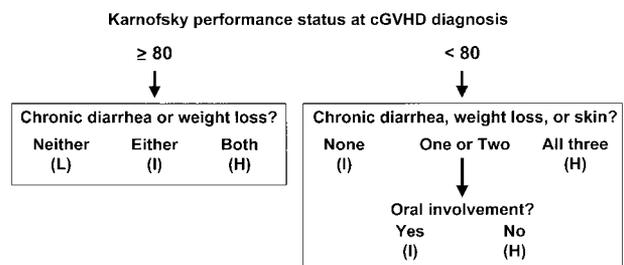


Figure 2. Chronic GVHD severity score.

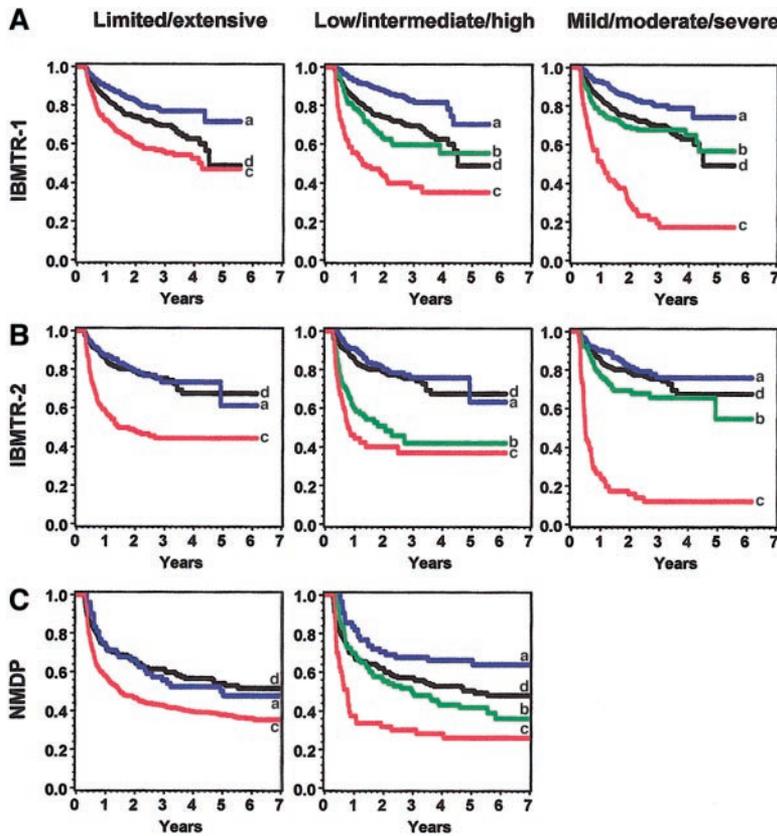


Figure 3. Overall survival in the various sets according to limited/extensive, low/intermediate/high risk, and mild/moderate/severe grading. Panel A shows IBMTR-1 training set; panel B, IBMTR-2 validation set; and panel C, NMDP validation set. Blue (a) indicates best risk; green (b), intermediate risk; red (c), worst risk; and black (d), no chronic GVHD.

**Description of cGVHD**

Cumulative incidence of cGVHD was greater (63% versus 42% at 1 year,  $P < .0001$ ), and manifestations more severe in the recipients of unrelated compared with sibling transplantations. Characteristics of cGVHD are shown in Table 2 and Figure 1. The greatest differences were seen in extensive disease (85% in recipients of unrelated compared with 46%-52% in sibling transplants,  $P < .0001$ ), lower KPS in recipients of unrelated transplants (median, 70% versus 80%,  $P < .0001$ ), and more frequent cutaneous involvement and diarrhea in recipients of unrelated transplants although less hepatic involvement was reported. However, median time to onset was similar in the 3 data sets (122-140 days,  $P =$  not significant). Skin, mouth, and liver involvement occurred in more

than one half of subjects with cGVHD, whereas eye involvement, diarrhea, and weight loss occurred in about one quarter (Figure 1). More than one half of subjects had 3 or more organs involved.

Incorrect designation of limited and extensive disease was common among recipients of sibling transplants with 65% to 67% of subjects scored as "limited," reporting organ involvement other than skin and liver. The most common other organ reported in patients misclassified as having limited disease was oral involvement (72%-83% of misclassified cases) and eye involvement (25%-41% of cases). Forty-three percent of the "limited" cGVHD in recipients of unrelated donor transplants was similarly misclassified, primarily because of oral involvement (83% of cases). However, the original classification as reported by the transplant

**Table 3. Overall survival**

Cohort	Any cGVHD		Lim, ext		Mild, mod, sev		Low, int, high	
	RR*	95% CI	RR*	95% CI	RR*	95% CI	RR*	95% CI
IBMTR-1	1.0	(0.8-1.2)	0.6	(0.4-0.8)	0.5	(0.4-0.7)	0.5	(0.3-0.7)
			1.3	(1.1-1.7)	1.0	(0.7-1.4)	1.2	(0.9-1.5)
IBMTR-2	1.4†	(1.0-1.8)	0.8	(0.5-1.2)	0.7	(0.4-1.0)	0.8	(0.5-1.2)
			2.0	(1.5-2.8)	1.1	(0.7-1.6)	2.0	(1.4-3.0)
NMDP	1.4‡	(1.1-1.9)	1.0	(0.7-1.6)	N/A		0.6§	(0.4-1.0)
			1.5	(1.2-2.0)			1.1§	(0.8-1.6)
						2.0	(1.3-3.0)	

Lim, ext indicates limited or extensive involvement; mild, mod, sev indicates mild, moderate, or severe clinical involvement; low, int, high indicates low risk, intermediate risk, high risk; RR, relative risk; CI, confidence interval; and N/A, not applicable.

\*Adjusted for disease type (acute myelogenous leukemia, acute lymphoblastic leukemia, chronic myelogenous leukemia), disease stage (early, intermediate, advanced), age (years), patient-donor sex matching, and presence of prior acute GVHD (yes, no).

† $P \leq .01$  compared with no cGVHD.

‡ $P \leq .001$  compared with no cGVHD.

§ $P \leq .01$  compared with the next most severe category.

|| $P \leq .0001$  compared with the next most severe category.

**Table 4. Treatment-related mortality**

Cohort	Any cGVHD		Lim, ext		Mild, mod, sev		Low, int, high	
	RR*	95% CI	RR*	95% CI	RR*	95% CI	RR*	95% CI
IBMTR-1	1.8†	(1.3-2.4)	0.7	(0.5-1.2)	0.6	(0.4-1.0)	0.6	(0.3-0.9)
			2.7	(2.0-3.7)	1.9	(1.4-2.8)	2.2§	(1.5-3.2)
IBMTR-2	2.8†	(1.9-4.1)	1.0	(0.6-1.9)	6.3	(4.4-9.0)	4.2	(2.9-6.0)
			4.5	(3.0-6.7)	1.0‡	(0.5-1.8)	1.4	(0.8-2.3)
NMDP	1.8†	(1.4-2.4)	1.4	(0.9-2.3)	2.1§	(1.3-3.5)	3.9	(2.4-6.4)
			1.9	(1.4-2.5)	10.9	(7.0-16.9)	5.2	(3.2-8.5)
				N/A		0.8	(0.5-1.5)	
						1.4‡	(0.9-2.1)	
						2.4	(1.5-3.7)	

For abbreviations see Table 3.

\*Adjusted for disease type (acute myelogenous leukemia, acute lymphoblastic leukemia, chronic myelogenous leukemia), disease stage (early, intermediate, advanced), age (years), patient-donor sex matching, and presence of prior acute GVHD (yes, no).

† $P \leq .0001$  compared with no cGVHD.

‡ $P \leq .01$  compared with next most severe category.

§ $P \leq .001$  compared with next most severe category.

|| $P \leq .0001$  compared with next most severe category.

centers was used for all subsequent analyses presented (ie, patients were not reclassified) because database limitations prevented our ability to apply the limited/extensive criteria with certainty.

**cGVHD severity score based on registry data**

Proportional hazards assumptions were fulfilled in the training set. Both statistical approaches identified similar clinical manifestations as important and generated identical prognostic groups. KPS, mouth and skin involvement, diarrhea, and weight loss were found to be important, with high KPS and mouth involvement being favorable prognostic signs and with skin involvement, diarrhea, and weight loss being unfavorable factors (Figure 1). Consideration of time to onset of cGVHD; platelet count; type of onset; eye, esophageal, liver, joint, and lung involvement at diagnosis did not improve prognostic ability.

Three prognostic groups were devised in which survival was statistically different. Low-risk patients had a high KPS ( $\geq 80\%$ ) and no diarrhea or weight loss. High-risk patients had high KPS, diarrhea, and weight loss or low KPS with either no oral involvement or all 3 poor prognostic signs: diarrhea, weight loss, and cutaneous involvement. Intermediate-risk patients had any KPS and all other combinations of organ involvement. The training population contained 50% low-, 32% intermediate-, and 18% high-risk patients (Figure 2).

There was no evidence of an interaction ( $P < .05$ ) between any of the variables associated with survival (disease type and disease

stage, age, donor-recipient sex-matching, and prior acute GVHD) and severity of cGVHD.

**Predictive ability of the cGVHD severity score in the validation sets**

The cGVHD severity score was applied to the validation sets after conversion of variables according to rules established before the analysis (Appendix). All analyses were adjusted for disease type and disease stage, recipient age, donor-recipient sex-matching, and prior acute GVHD. Survival as predicted by different scoring systems is shown in Figure 3 and summarized in Table 3. Relative risks (RRs) represent the effect of cGVHD severity on survival compared with the baseline group of patients without cGVHD. When the new severity score was applied to the IBMTR-2 validation set, low-risk patients had better survival than the intermediate- and high-risk group (RR, 0.8 versus 2.0 versus 2.7,  $P < .0001$ ); intermediate- and high-risk groups were similar ( $P = .29$ ). In the NMDP validation set, patients with low-, intermediate-, and high-risk cGVHD had statistically different survival times (RR, 0.6 versus 1.1 versus 2.0,  $P = .0006$ ).

**Comparison of grading schemes**

Comprehensive comparison of the low-/intermediate-/high-risk schema, the limited/extensive score, and the IBMTR mild/moderate/severe

**Table 5. Relapse**

Cohort	Any cGVHD		Lim, ext		Mild, mod, sev		Low, int, high	
	RR*	95% CI	RR*	95% CI	RR*	95% CI	RR*	95% CI
IBMTR-1	0.5‡	(0.3-0.7)	0.5	(0.3-0.8)	0.5	(0.3-0.7)	0.5	(0.3-0.8)
			0.4	(0.3-0.7)	0.3	(0.2-0.6)	0.4	(0.2-0.7)
IBMTR-2	0.6†	(0.4-0.9)	0.8§	(0.5-1.4)	0.7	(0.4-1.5)	0.4	(0.2-0.9)
			0.3	(0.2-0.7)	0.6	(0.4-1.1)	0.5	(0.3-0.9)
NMDP	0.6†	(0.3-1.0)	0.5	(0.2-1.1)	0.6	(0.3-1.1)	1.0	(0.5-1.9)
			0.6	(0.3-1.1)	0.3	(0.1-1.4)	0.3	(0.1-1.0)
				N/A		0.4	(0.2-1.2)	
						0.7	(0.3-1.5)	
						1.0	(0.3-2.9)	

For abbreviations see Table 3.

\*Adjusted for disease type (acute myelogenous leukemia, acute lymphoblastic leukemia, chronic myelogenous leukemia), disease stage (early, intermediate, advanced), age (years), patient-donor sex matching, and presence of prior acute GVHD (yes, no).

† $P \leq .01$  compared with no cGVHD.

‡ $P \leq .0001$  compared with no cGVHD.

§ $P \leq .01$  compared with the next most severe category.

**Table 6. Disease-free survival**

Cohort	Any cGVHD		Lim, ext		Mild, mod, sev		Low, int, high	
	RR*	95% CI	RR*	95% CI	RR*	95% CI	RR*	95% CI
IBMTR-1	1.0	(0.8-1.2)	0.6¶	(0.4-0.8)	0.5	(0.4-0.7)	0.5¶	(0.4-0.7)
			1.3	(1.0-1.6)	1.0¶	(0.7-1.3)	1.1¶	(0.8-1.4)
IBMTR-2	1.3†	(1.0-1.7)	0.9	(0.6-1.3)	2.9	(2.2-3.9)	2.0	(1.5-2.7)
			1.8	(1.3-2.4)	0.8	(0.5-1.1)	0.8¶	(0.6-1.2)
NMDP	1.5‡	(1.1-1.9)	1.2	(0.8-1.7)	4.5	(3.1-6.5)	2.2	(1.5-3.3)
			1.5	(1.2-2.0)	N/A		0.7	(0.4-1.2)
						1.2§	(0.8-1.7)	
						2.0	(1.3-3.1)	

For abbreviations see Table 3.

\*Adjusted for disease type (acute myelogenous leukemia, acute lymphoblastic leukemia, chronic myelogenous leukemia), disease stage (early, intermediate, advanced), age (years), patient-donor sex matching, and presence of prior acute GVHD (yes, no).

† $P \leq .01$  compared with no cGVHD.

‡ $P \leq .001$  compared with no cGVHD.

§ $P \leq .01$  compared with the next most severe category.

|| $P \leq .001$  compared with the next most severe category.

¶ $P \leq .0001$  compared with the next most severe category.

clinical grading was hampered by incorrect classification of the limited/extensive scale by transplant centers and absence of clinical impression for NMDP subjects. However, the new severity score was equivalent or better than the limited/extensive classification based on lower AIC, but not as good as the mild/moderate/severe classification in predicting survival (Figure 3). All grading schemes performed best in the IBMTR-1 cohort. The limited/extensive analysis was based on data reported directly by the centers; no patients were reclassified. When patients were reclassified by using reported organ involvement according to our algorithm, predictive ability worsened in all 3 cohorts.

#### Effects of cGVHD severity on treatment-related mortality, relapse, and disease-free survival

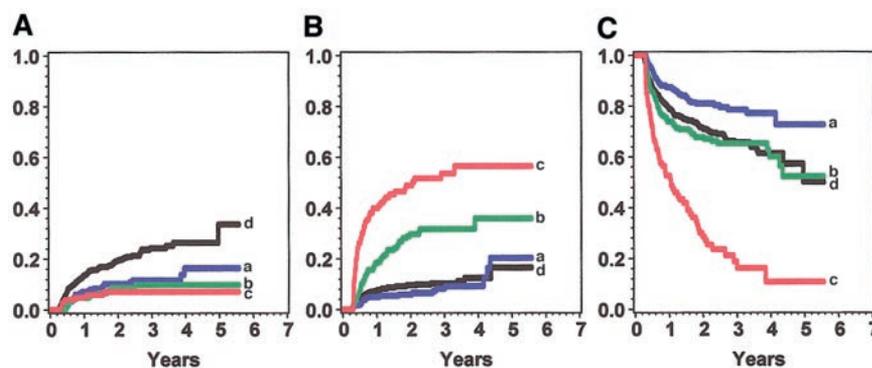
The effects of cGVHD severity on treatment-related mortality, relapse, and disease-free survival are shown in Tables 4, 5, and 6 and graphed for the IBMTR-1 cohort in Figure 4. In all 3 cohorts, treatment-related mortality increased with increasing severity of cGVHD with the use of all 3 grading systems. For example, in the IBMTR-1 cohort, intermediate risk (RR, 2.2; 95% confidence interval [CI], 1.5-3.2) and high risk (RR, 4.2; 95% CI, 2.9-6.0) were associated with more treatment-related mortality than low risk or no cGVHD. A similar pattern was seen in the IBMTR-2 and NMDP groups.

A lower relapse rate was seen in all patients with cGVHD, but we could find no evidence that relapse rate was associated with cGVHD severity. Relapses were uncommon after the development of cGVHD (8%-9%), and increasing severity of cGVHD was not associated with a longer time to relapse.

## Discussion

Prior reports have noted that patients with cGVHD have fewer leukemia relapses<sup>30-32</sup> but more treatment-related mortality when considered as a binary end point.<sup>9,10</sup> We confirm and extend these findings by considering cGVHD severity. By using 3 grading systems: limited/extensive, a clinical impression scale (mild, moderate, severe), and a new severity score developed and validated specifically for this study, we could not find evidence that severity of cGVHD was associated with relapse risk (RR, 0.5-0.6), but we did find that more severe cGVHD is associated with more treatment-related mortality. The net effect of these 2 influences determines the effect of cGVHD on disease-free and overall survival. Greater cGVHD is associated with worse survival because there is greater treatment-related mortality without a benefit of fewer relapses.

We based our severity score on data reported retrospectively to the IBMTR and NMDP. This score separates subjects into low-, intermediate-, and high-risk categories based on KPS at diagnosis and on the presence of diarrhea, weight loss, cutaneous manifestations, and/or oral involvement. For example, patients in the low-risk group comprised 50% of the IBMTR-1 cGVHD population, 61% of the IBMTR-2 group, and 34% of the NMDP patients. They had high KPS ( $\geq 80\%$ ) and no diarrhea or weight loss. The new score performed better than the limited/extensive grading system in predicting overall and disease-free survival but not as well as the clinical impression of mild/moderate/severe disease. However, within all 3 grading schemes we could identify a favorable subgroup of patients (low risk, limited



**Figure 4. Condition according to the new severity score in the IBMTR-1 cohort.** Panel A shows relapse; panel B, treatment-related mortality; and panel C, disease-free survival. Blue (a) indicates low risk; green (b) intermediate risk; red (c) high risk; and black (d), no chronic GVHD.

involvement, mild severity) whose survival was as good as, or better than, patients without cGVHD because of low treatment-related mortality and high antileukemic effect.

Several factors associated with poor prognosis of patients with cGVHD are reported by others, including lichenoid skin changes (RR, 2.2 and 2.5), skin involvement on more than 50% of body surface area (RR, 7.0), elevated total bilirubin (RR, 2.1), progressive onset (RR, 1.7 and 4.1), low platelet count (RR, 2.0 and 3.6), and cGVHD developing after corticosteroid refractory/dependent acute GVHD.<sup>6,8,25,26</sup> However, many of these variables are not available in our databases (lichenoid skin involvement, extent of skin involvement, bilirubin levels, prior corticosteroid treatment) or were not found to have prognostic significance (progressive onset, low platelet counts) in these analyses.

An ideal grading scheme has the following properties: (1) predicts survival or disease-free survival in a clinically meaningful way even as transplantation procedures evolve; (2) applicable in diverse populations, including related and unrelated transplants, HLA-matched and mismatched recipients, and both children and adults; (3) reproducible over time and at different sites; and, finally, (4) easy to apply. Each grading scheme we used in this analysis suffered from limitations in one or more of these areas.

The mild/moderate/severe clinical impression appeared to be the most predictive but had 2 major limitations. First, centers classified their patients on the basis of clinical impression without specific guidance as to these definitions. In the absence of objective criteria, reproducibility, reliability, and validity of individual assessments are questionable. Second, these classifications may reflect a lengthy period of observation. The designation of mild/moderate/severe cGVHD may be highly predictive because it incorporates response to treatment, evolution of new organ involvement, and even knowledge about morbidity and death as a result of cGVHD. Only with real-time data defining the worst manifestations of cGVHD at a given time point can these ambiguities be eliminated.

The limited/extensive grading scheme appears relatively straightforward, but our results show application in practice can be

difficult. Although the criteria are printed on the data forms, many clinical situations do not fit either the limited or extensive category, and many centers appear to misclassify patients based on inconsistencies in the spectrum of organs reported. In patients without clinical or histologic characteristics detailed by limited or extensive headings, physicians may interpret the scale as an indicator of severity. If so, the limited/extensive scale becomes a subjective assessment similar to the mild/moderate/severe system.

Our severity score performed well in the training set, but it is difficult to use. Also, it performed less well in the validation sets. This finding demonstrates the need to validate proposed scales either prospectively or in independent data sets with a fuller representation of all relevant data elements. Although registry data can indicate potentially important components of a severity scale, we suggest that any further refinement in cGVHD assessment is best accomplished by a prospective effort to collect detailed information on characteristics at cGVHD onset, treatment response, and course of organ involvement in a large group of patients by using defined criteria.

Despite these caveats, our analysis shows that patients identified by centers as having extensive cGVHD, severe involvement, or high-risk disease are likely to have significantly worse survival than patients without cGVHD or with less severe cGVHD. These patients are appropriate candidates for clinical trials aimed at improving management of cGVHD. Patients with limited cGVHD, mild involvement, or low-risk as defined by our criteria (KPS  $\geq$  80%, no diarrhea, no weight loss) have survival equal to or better than patients without cGVHD and may not need aggressive treatment or an extended duration of immune suppression.

### Acknowledgment

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

## Appendix

### Compatibility of cGVHD variables from the 3 data sets

	IBMTR-1 training	IBMTR-2 validation	NMDP validation
N	1827	1092	553
Years of transplant	1990-1994	1990-1997	1990-1994
Version of data collection form	089	095	NMDP
Cutaneous*	Mild, moderate, or severe cutaneous involvement	Subclinical (biopsy only), rash, scleroderma, dyspigmentation, contractures, alopecia, other skin involvement present	Cutaneous involvement, contractures present
Ocular	Mild, moderate, or severe ocular involvement	Dry eyes, corneal erosions/conjunctivitis, other eye involvement present	Xerophthalmia (dry eyes) present
Oral*	Mild, moderate, or severe oral involvement	Lichenoid changes, mucositis/ulcers, other mouth involvement present	Oral involvement, mucositis present
Diarrhea*	Mild, moderate, or severe chronic diarrhea	Chronic diarrhea present	Chronic diarrhea present
Esophagus	Mild, moderate, or severe esophageal involvement	Esophageal involvement present	Esophageal involvement present
Hepatitis	Mild, moderate, or severe hepatitis	Liver involvement present	Hepatitis/hepatic involvement present
Joint	Mild, moderate, or severe joint involvement	Joint involvement present	Arthritis/arthralgia (joint pain) present
Pulmonary	Mild, moderate, or severe obstructive lung disease	Bronchiolitis obliterans or other lung involvement present	Obstructive lung disease present
Weight loss*	Mild, moderate, or severe weight loss	Chronic nausea/vomiting or malabsorption present (weight loss not specifically collected)	Weight loss present

\*Used in final severity score.

## References

- 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;75:2459-2464.
- Ochs LA, Miller WJ, Filipovich AH, et al. Predictive factors for chronic graft-versus-host disease after histocompatible sibling donor bone marrow transplantation. *Bone Marrow Transplant*. 1994;13:455-460.
- Sutherland HJ, Fyles GM, Adams G, et al. Quality of life following bone marrow transplantation: a comparison of patient reports with population norms. *Bone Marrow Transplant*. 1997;19:1129-1136.
- Syrjala KL, Chapko MK, Vitaliano PP, Cummings C, Sullivan KM. Recovery after allogeneic marrow transplantation: prospective study of predictors of long-term physical and psychosocial functioning. *Bone Marrow Transplant*. 1993;11:319-327.
- Duell T, van Lint MT, Ljungman P, et al. Health and functional status of long-term survivors of bone marrow transplantation: EBMT Working Party on Late Effects and EULEP Study Group on Late Effects: European Group for Blood and Marrow Transplantation. *Ann Intern Med*. 1997;126:184-192.
- Sullivan KM, Witherspoon RP, Storb R, et al. Prednisone and azathioprine compared with prednisone and placebo for treatment of chronic graft-v-host disease: prognostic influence of prolonged thrombocytopenia after allogeneic marrow transplantation. *Blood*. 1988;72:546-554.
- Sullivan KM, Witherspoon RP, Storb R, et al. Alternating-day cyclosporine and prednisone for treatment of high-risk chronic graft-v-host disease. *Blood*. 1988;72:555-561.
- 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.
- Loughran TP Jr, Sullivan K, Morton T, et al. Value of day 100 screening studies for predicting the development of chronic graft-versus-host disease after allogeneic bone marrow transplantation. *Blood*. 1990;76:228-234.
- 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.
- Majolino I, Saglio G, Scime R, et al. High incidence of chronic GVHD after primary allogeneic peripheral blood stem cell transplantation in patients with hematologic malignancies. *Bone Marrow Transplant*. 1996;17:555-560.
- Storek J, Gooley T, Siadak M, et al. Allogeneic peripheral blood stem cell transplantation may be associated with a high risk of chronic graft-versus-host disease. *Blood*. 1997;90:4705-4709.
- Urbano-Ispizua A, Garcia-Conde J, Brunet S, et al. High incidence of chronic graft versus host disease after allogeneic peripheral blood progenitor cell transplantation: The Spanish Group of Allo-PBPCT. *Haematologica*. 1997;82:683-689.
- Solano C, Martinez C, Brunet S, et al. Chronic graft-versus-host disease after allogeneic peripheral blood progenitor cell or bone marrow transplantation from matched related donors: a case-control study: Spanish Group of Allo-PBT. *Bone Marrow Transplant*. 1998;22:1129-1135.
- Scott MA, Gandhi MK, Justice HK, Mahendra P, Bass G, Marcus RE. A trend towards an increased incidence of chronic graft-versus-host disease following allogeneic peripheral blood progenitor cell transplantation: a case controlled study. *Bone Marrow Transplant*. 1998;22:273-276.
- Ustun C, Arslan O, Beksac M, et al. A retrospective comparison of allogeneic peripheral blood stem cell and bone marrow transplantation results from a single center: a focus on the incidence of graft-vs.-host disease and relapse. *Biol Blood Marrow Transplant*. 1999;5:28-35.
- Blaise D, Kuentz M, Fortanier C, et al. Randomized trial of bone marrow versus lenograstim-primed blood cell allogeneic transplantation in patients with early-stage leukemia: a report from the Societe Francaise de Greffe de Moelle. *J Clin Oncol*. 2000;18:537-546.
- Weiden PL, Sullivan KM, Flournoy N, Storb R, Thomas ED. Antileukemic effect of chronic graft-versus-host disease: contribution to improved survival after allogeneic marrow transplantation. *N Engl J Med*. 1981;304:1529-1533.
- Sullivan KM, Weiden PL, Storb R, et al. Influence of acute and chronic graft-versus-host disease on relapse and survival after bone marrow transplantation from HLA-identical siblings as treatment of acute and chronic leukemia [published erratum appears in *Blood*. 1989;74:1180]. *Blood*. 1989;73:1720-1728.
- Horowitz MM, Gale RP, Sondel PM, et al. Graft-versus-leukemia reactions after bone marrow transplantation. *Blood*. 1990;75:555-562.
- Ringden O, Labopin M, Gluckman E, et al. Strong antileukemic effect of chronic graft-versus-host disease in allogeneic marrow transplant recipients having acute leukemia treated with methotrexate and cyclosporine: The Acute Leukemia Working Party of the European Group for Blood and Marrow Transplantation (EBMT). *Transplant Proc*. 1997;29:733-734.
- Gratwohl A, Hermans J, Apperley J, et al. Acute graft-versus-host disease: grade and outcome in patients with chronic myelogenous leukemia: Working Party Chronic Leukemia of the European Group for Blood and Marrow Transplantation. *Blood*. 1995;86:813-818.
- Gratwohl A, Brand R, Apperley J, et al. Prognostic relevance of acute graft-versus-host disease (AGVHD) on outcome in patients transplanted with HLA-identical sibling transplants for chronic myeloid leukemia (CML) in first chronic phase [abstract]. *Blood*. 2000;96:193a.
- 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.
- Morton A, Anasetti C, Gooley T, et al. Chronic graft-versus-host disease (GVHD) following unrelated donor transplantation. *Blood*. 1997;90:590a.
- 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.
- Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov EBN, Csaki F, eds. 2nd International Symposium of Information Theory and Control. Budapest, Hungary: Akademia Kiado; 1973; 267-281.
- Klein J, Moeschberger M. *Survival Analysis: Techniques for Censored and Truncated Data*. New York, NY: Springer-Verlag; 1997.
- Pepe MS, Longton G, Pettinger M, Mori M, Fisher LD, Storb R. Summarizing data on survival, relapse, and chronic graft-versus-host disease after bone marrow transplantation: motivation for and description of new methods. *Br J Haematol*. 1993;83:602-607.
- Goldman JM, Gale RP, Horowitz MM, et al. Bone marrow transplantation for chronic myelogenous leukemia in chronic phase: increased risk for relapse associated with T-cell depletion. *Ann Intern Med*. 1988;108:806-814.
- Enright H, Davies SM, DeFor T, et al. Relapse after non-T-cell-depleted allogeneic bone marrow transplantation for chronic myelogenous leukemia: early transplantation, use of an unrelated donor, and chronic graft-versus-host disease are protective. *Blood*. 1996;88:714-720.
- van Rhee F, Szydlo RM, Hermans J, et al. Long-term results after allogeneic bone marrow transplantation for chronic myelogenous leukemia in chronic phase: a report from the Chronic Leukemia Working Party of the European Group for Blood and Marrow Transplantation. *Bone Marrow Transplant*. 1997;20:553-560.