Biol Blood Marrow Transplant 25 (2019) 1761–1769

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

Combination of the Centre for International Blood and Marrow Transplant Registry Risk Score and the Global Severity Score Enhances Prognostic Risk Stratification in Patients Receiving Frontline Therapy for Chronic Graft-versus-Host Disease

Roman M. Shapiro[1], Elizabeth Shin[2], Arjun Datt Law[3], Wilson Lam[3], Fotios V. Michelis[3], Auro Viswabandya[3], Rajat Kumar[3], Jeffrey H. Lipton[3], Hans Messner[3], Jonas Mattsson[3,4], Dennis Dong Hwan Kim[3,*]

[1] *Adult Hematology Program, University of Toronto, Toronto, Ontario, Canada*
[2] *Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada*
[3] *Allogeneic Blood and Marrow Transplant Program, Princess Margaret Cancer Centre, University of Toronto, Toronto, Ontario, Canada*
[4] *Department of Oncology and Pathology, Karolinska Institute, Stockholm, Sweden*

Article history:
Received 5 March 2019
Accepted 26 May 2019

Key words:
Chronic
GVHD
Prognostic
Risk
Score
Survival

A B S T R A C T

The Centre for International Blood and Marrow Transplant Registry (CIBMTR) score has been shown to be prognostic for overall survival (OS) and nonrelapse mortality (NRM) but has been shown in several single-center studies to classify a large proportion of patients with chronic graft-versus-host disease (cGVHD) in the lower risk groups (RG1 to RG2), thereby limiting its prognostic utility for those patients. We evaluate the CIBMTR score, the Global Severity Score (GSS), and a novel risk score developed to improve on the limitations of the CIBMTR with respect to clinically relevant outcomes, including failure-free survival (FFS), in patients receiving frontline systemic treatment for cGVHD. We identified 277 patients between 2002 and 2012 at the Princess Margaret Cancer Centre in Toronto, Canada, who developed cGVHD and were treated with at least 1 line of systemic therapy. cGVHD was graded by GSS, and patients were stratified by CIBMTR. We evaluated OS, NRM, relapse, and FFS within GSS grade groups, as well as CIBMTR RGs, and used a novel prognostic risk score. The median FFS duration was 164 days in the severe GSS group versus 238 days in the moderate-grade group and 304 days in mild-grade group (P= .001). The median FFS duration was 501 days in CIBMTR RG1 versus 291 days in RG2 and 166 days in RG3 to RG6 (P = .003). A novel risk score combining the GSS and CIBMTR scores was prognostic of OS, NRM, and FFS and was able to subdivide patients with cGVHD in CIBMTR RG1 to RG2 into distinct prognostic risk categories. The CIBMTR risk score and the GSS are well correlated with FFS, OS, and NRM following frontline systemic treatment for cGVHD. A new risk score model combining the CIBMTR risk score and the GSS could enhance risk stratification in the lower CIBMTR risk groups.

INTRODUCTION

Chronic graft-versus-host disease (cGVHD) is one of the major complications frequently occurring after allogeneic hematopoietic stem cell transplantation (HSCT). Prognostic risk stratification of patients who develop cGVHD serves as an important factor clinically when discussing anticipated outcomes [1]. Numerous prognostic risk scores had been developed for cGVHD, including the National Institutes of Health (NIH) consensus criteria (NCC) for the diagnosis of cGVHD in 2005 [2]. Since the introduction of the NCC, 2 prognostic risk systems have been widely used in clinical practice that have been validated in the setting of the newer diagnostic criteria for cGVHD [3–5].

The Centre for International Blood and Marrow Transplant Registry (CIBMTR) cGVHD risk score had been originally introduced for prognostication in patients diagnosed with cGVHD based on the revised Seattle criteria [6]. A large multicenter cohort of 5343 patients with cGVHD diagnosed based on the revised Seattle criteria demonstrated that the use of 10 clinical variables (5 pretransplant variables and 5 post-HSCT variables) could evenly stratify patients with cGVHD according to overall survival (OS) and nonrelapse mortality (NRM) outcomes after

Financial disclosure: See Acknowledgments on page 1769.

* Correspondence and reprint requests: Dennis (Dong Hwan) Kim, MD, PhD, Allogeneic Blood and Marrow Transplant Program, Department of Medical Oncology and Hematology, Princess Margaret Cancer Centre, University of Toronto, 610 University Avenue, Toronto, Ontario, Canada M5G2M9.
E-mail address: dr.dennis.kim@uhn.ca (D.D.H. Kim).

https://doi.org/10.1016/j.bbmt.2019.05.029
1083-8791/

initiation of cGVHD treatment [3]. The 10 variables included in this risk score are recipient age, donor type, sex mismatch, disease status before HSCT, GVHD prophylaxis, prior acute graft-versus-host disease (GVHD), time from transplantation to cGVHD, platelet count, serum bilirubin, and performance status at the onset of cGVHD [4]. More recently, this CIBMTR risk score was successfully validated at 2 individual centers, the Fred Hutchinson Cancer Research Centre and the Princess Margaret Cancer Centre, in patients with leukemia or myelodysplastic syndrome who received HSCT between 2006 and 2010 and developed cGVHD based on the NCC [4,6]. In these validation cohorts, the CIBMTR risk score performed well in predicting differences in OS and was confirmed as a useful tool to stratify the risk of mortality at the onset of cGVHD.

An important limitation of the CIBMTR score, however, is that it classifies most patients into the lower risk groups [4,7]. In the validation cohort at the Princess Margaret Cancer Centre with patients with myelodysplastic syndrome and acute leukemia, for example, nearly 80% were classified into CIBMTR risk group (RG) 2 [7]. This classification indicates a need for improvement of the prognostic RGs generated by the CIBMTR, especially for most patients who fall into the same RG.

The other risk score widely used for cGVHD in clinical practice is the Global Severity Score (GSS), initially developed as an assessment tool of cGVHD severity [5,8]. The GSS differs from the CIBMTR risk score in that the latter does not take into account actual disease severity based on organ damage and functional impairment. It stratifies disease severity as mild, moderate, or severe based on degree of organ involvement and number of organs affected. Several studies have validated the use of the GSS as a prognostic tool in patients with cGVHD using the OS as the clinical endpoint [9,10].

Given the GSS evaluates aspects of cGVHD not assessed by the CIBMTR, it is conceivable to hypothesize that some combination of the GSS and the CIBMTR risk scores may yield an improved risk score that yields additional prognostic information for patients within the predominant CIBMTR RG. Assessment of such a combination in terms of its ability to generate prognostic groups with respect to clinically relevant outcomes such as OS and NRM is one of the major goals of this study.

The other major goal of this study is to validate the CIBMTR and GSS prognostic risk scores as well as any newly developed risk score with respect to another clinical outcome measure, the failure-free survival (FFS). FFS is an intermediate endpoint that has been proposed for prognostication of patients with cGVHD with the benefit of being evaluable at a shorter time point than OS [7,11,12]. The definition of FFS includes switch of treatment caused by failure, disease relapse, and NRM. Implementation of FFS into prospective clinical trial design would be very useful because it would allow for the evaluation of cGVHD with a shorter follow-up duration than OS [13]. Such an implementation would be facilitated if the classification of patients into distinct prognostic groups by the GSS, the CIBMTR, or any other novel risk scores could be validated with respect to this relevant clinical outcome.

PATIENTS AND METHODS

We retrospectively reviewed 668 consecutive patients who underwent allogeneic HSCT between 2002 and 2012 at the Princess Margaret Cancer Centre, Toronto, Canada. We then identified 312 patients who had a diagnosis of GVHD, of whom 277 had a diagnosis of cGVHD and received at least 1 systemic treatment for this. The late-onset acute GVHD was excluded from the analysis. In the final analysis, 277 patients were included who received a prednisone-based regimen for first-line treatment for cGVHD. The research ethics board at University Health Network approved this study.

Transplant Procedures and GVHD Treatment

Unrelated donors of peripheral blood stem cells or bone marrow cells were identified through the OneMatch Stem Cell and Marrow Donor Network. Patients received conditioning regimens before HSCT infusion as per institutional protocols. GVHD prophylaxis consisted of cyclosporine (administered for 3 months in related donor transplants unless GVHD required further treatment and for 6 months in unrelated donor transplants), combined with either methotrexate (15 mg/m^2 on day +1 and 10 mg/m^2 on days +3, +6, and +11 following allogeneic HSCT) or mycophenolate mofetil (15 mg/kg every 8 hours until day +30).

The initial treatment for acute GVHD and cGVHD included oral prednisone or intravenous methylprednisolone, with the starting dose chosen at the discretion of the treating physician. When additional treatment was indicated in the setting of cGVHD because of failure of initial therapy for reasons such as refractory GVHD, a flare of GVHD, or intolerance to previous lines of immunosuppression (IST), second or further lines of IST were introduced according to each patient's clinical manifestations or toxicity profiles at the physician's discretion. These salvage regimens included azathioprine, mycophenolate mofetil, tacrolimus, extracorporeal photopheresis, rituximab, or others.

Definitions

Chronic GVHD was classified and graded using the 2005 NIH consensus criteria. Overlap cGVHD referred to patients who had diagnostic features of cGVHD and who also met any criteria for acute GVHD. The GSS was evaluated in patients with cGVHD individually at the time of initiation of systemic therapy and graded as mild, moderate, or severe. The CIBMTR risk score was also calculated at the time of initiation of systemic therapy for cGVHD as previously described [3]. Variable-specific risk scores were summed for each patient to assign an overall risk score. RGs were assigned based on the overall risk score as follows: RG1, 0 to 2; RG2, 3 to 6; RG3, 7 to 8; RG4, 9 to 10; RG5, 11; and RG6, 12.

Treatment failure was defined as the initiation of the next line of IST for cGVHD or a re-escalation of the dose of prednisone to $\geq$1 mg/kg/d regardless of the target organs. OS was calculated from the date of initiation of systemic therapy for cGVHD to the date of death or last follow-up. FFS was defined as the time period between the date of initiation of systemic therapy of cGVHD to the date of treatment failure, NRM, or relapse of disease.

Statistical Analysis

We evaluated OS, NRM, and FFS after initiation of frontline cGVHD treatment. The OS and FFS were calculated using the Kaplan-Meier method with the log-rank test for univariate analysis. The cumulative incidences of NRM, disease relapse, and treatment failure were estimated considering competing risks using Gray's method for univariate analysis. Disease relapse, NRM, and treatment failure were competing risks for each other.

The Cox proportional hazard regression model was used for FFS and OS, whereas Fine-Gray methods were used for the incidences of NRM and relapse. We decided to include 4 risk factors into the multivariate model, including the RG by the CIBMTR risk score, the GSS by the NCC, cGVHD subtype (classical versus overlap syndrome), and source of stem cells. Factors with a P value $<$.1 in the univariate analyses were entered in the multivariate analysis. A stepwise selection algorithm was applied for model selection using the criteria for variable selection, P = .1 for variable entry and P = .05 for variable removal.

Next, a risk score model was generated using the RG of the CIBMTR risk score and the GSS by the NCC identified as independent risk factors for FFS, OS, and NRM consistently. For the CIBMTR risk score, score 0 was assigned to RG1 and RG2 (ie, CIBMTR risk score 0 to 6), whereas score 1 was assigned to RG3 to RG6 (ie, CIBMTR risk score 7 to 12). For the GSS by the NCC, score 0 was assigned to mild- or moderate-grade cGVHD, whereas score 1 was assigned to severe-grade cGVHD. A risk score was summed from both scores of CIBMTR risk score and GSS by the NCC. Risk scores of 2 were classified as high risk, scores of 1 were intermediate risk, and a score of 0 was low risk. The FFS, OS, and NRM were compared according to the new risk score.

To formally compare the new risk score with the CIBMTR in terms of risk-stratifying patients with respect to the clinical outcomes of FFS, NRM, and OS, we performed the net reclassification improvement (NRI) test. This is an appropriate test to quantify any possible improvement in the new risk model in comparison with the CIBMTR because the new model is constructed with the addition of the NIH GSS to the CIBMTR score [14]. The NRI implements a risk difference cutoff for the difference in probability between being correctly upstaged and incorrectly downstaged in the case of obtaining an event for a clinical outcome. We set the risk difference cutoff to determine a risk difference between the 2 scores of at least 5%.

The statistical analysis for cumulative incidence and survival analyses as well as NRI were performed using EZR software (Saitama Medical Centre, Jichi Medical University, Saitama, Japan). EZR is a modified version of R commander (version 2.12.1) [15].

RESULTS

Patient/Transplant Characteristics, CIBMTR cGVHD Risk Score, and GSS by the NIH Consensus Criteria

Of 668 patients who received HSCT between 2002 and 2012 at the Princess Margaret Cancer Centre, 277 were diagnosed with cGVHD at a median onset of 140 days (range, 45 to 381 days) based on the NCC and received prednisone-based first-line treatment. In total, 102 patients (36.9%) were diagnosed with classical cGVHD and 175 patients (63.1%) with overlap syndrome. cGVHD was classified by the GSS score as mild in 90 patients (32.5%), moderate in 143 patients (51.6%), and severe in 44 patients (15.9%) at the onset of cGVHD (Table 1). Thirty-three patients (11.9%) had progressive-type onset. According to the CIBMTR risk score, of 227 patients with available data, 32 were in RG1 (score 0 to 2; 14.1%), 162 patients in RG2 (score 3 to 6; 71.4%), 27 patients in RG3 (score 7 to 8; 11.9%), 5 patients in RG4 (score 9 to 10; 2.2%), and 1 patient in RG5 (score 11; 0.4%) (Table 1).

FFS after Frontline cGVHD Treatment

With a median follow-up duration of 26 months among survivors, the median duration of FFS was 255 days (95% confidence interval [CI], 219 to 321 days). The FFS rate was 63.8% (95% CI, 56.2% to 67.8%) at 6 months, 38.4% (95% CI, 32.4% to 44.3%) at 12 months, and 27.1% (95% CI, 21.5% to 33.0%) at 24 months, respectively (Figure 1). In the calculation of FFS, the most common cause of treatment failure was switch of immunosuppressive therapy. The probability of OS at 2 years was 73.5% (95% CI, 67.0% to 78.8%), whereas the cumulative incidence of NRM was 17.1% (95% CI, 12.4% to 22.2%) at 2 years after onset of cGVHD.

Both the CIBMTR and GSS by the NCC risk scores are able to stratify patients well based on both FFS and OS (Figure 2). The group with higher CIBMTR risk score showed a shorter FFS duration: 166 days (95% CI, 94 to 237 days) in RG3 to RG6 versus 291 days (95% CI, 216 to 365 days) in RG2 versus 501 days (95% CI, 0 to 1083 days) in RG1 ($P = .003$). The group with the severe grade of GSS had a shorter FFS compared with others: median FFS duration of 164 days (95% CI, 122 to 205 days) in the severe-grade versus 238 days (95% CI, 153 to 340 days) in the moderate-grade versus 304 days (95% CI, 237 to 371 days) in the mild-grade group ($P = .001$).

OS, NRM, and Relapse Incidence According to the CIBMTR cGVHD Risk Score and the GSS by the NCC Risk Score

The group with a high CIBMTR risk score showed significantly worse OS and a higher risk of NRM (Table 2). The OS rate at 2 years was 92.0% (95% CI, 71.5% to 98%) in RG1 versus 81.5% (95% CI, 73.3% to 87.4%) in RG2 versus 38.9% (95% CI, 21.2% to 56.2%) in RG3 to RG6 ($P < .001$), whereas the NRM rate at 2 years was 3.7% (95% CI, 0.3% to 16.2%) in RG1 versus 13.3% (95% CI, 8.0% to 20.0%) in RG2 versus 51.3% (95% CI, 31.0% to 68.3%) in RG3 to RG6 ($P < .001$). Similarly, the group with severe-grade GSS also showed significantly worse OS and a higher NRM rate (Table 2). The OS rate at 2 years was 74.5% (95% CI, 62.9% to 83.0%) in the mild-grade versus 79.5% (95% CI, 70.1% to 86.2%) in the moderate-grade and 55.0% (95% CI, 38.7% to 68.6%) in the severe-grade GSS group ($P < .001$), whereas the NRM rate at 2 years was 10.2% (95% CI, 4.7% to 18.2%) in the mild-grade versus 13.7% (95% CI, 7.9% to 21.1%) in the moderate-grade and 40.3% (95% CI, 25.3% to 54.9%) in the severe-grade group ($P < .001$).

Multivariate Analysis for FFS, OS, and NRM

Having shown a significant correlation between the CIBMTR and GSS risk scores and the outcomes of FFS, OS, and NRM in

Table 1
Patients and Transplant Characteristics (n = 277)

Variable	No. of Patients	%
Age		
Years, median (range)	51 (19-70)	
<30 years	30	10.8
30-59 years	189	68.2
≥60 years	58	20.9
Sex		
Male	162	58.5
Female	115	41.5
Sex match		
Female to male	65	23.5
Others	212	76.5
Underlying diagnosis		
AML	120	43.3
ALL	23	8.3
MDS	27	9.7
CML	20	7.2
MPN	23	8.3
CLL	24	8.7
Lymphoma	33	11.9
Others	7	2.6
Conditioning intensity		
Myeloablative	180	65.0
Reduced intensity	97	35.0
HLA and donor type		
Matched sibling	175	63.2
Matched unrelated	82	29.6
Mismatched unrelated	20	7.2
Stem cell source		
Bone marrow	20	7.2
Peripheral blood	257	92.8
GVHD prophylaxis		
CNI + MTX	82	29.6
CNI + MMF	141	50.9
CNI + TCD	37	13.4
Others	17	6.1
Onset of cGVHD		
Days, median (range)	140 (45-381)	
<5 months	154	55.6
≥5 months	123	44.4
cGVHD subtype		
Classical cGVHD	102	36.9
Overlap syndrome	175	63.1
cGVHD global score		
Mild grade	90	32.5
Moderate grade	143	51.6
Severe grade	44	15.9
CIBMTR cGVHD risk score (n = 227 available)		
RG1 (score 0-2)	32	14.1
RG2 (score 3-6)	162	71.4
RG3 (score 7-8)	27	11.9
RG4 (score 9-10)	5	2.2
RG5 (score 11)	1	0.4
RG6 (score 12)	0	0
Organ involvement at initial presentation		
Eye	103	37.2
Mouth	147	53.1
Skin	189	68.2

(*continued*)

Table 1 (*Continued*)

Variable	No. of Patients	%
GIT	64	23.1
Liver	208	75.1
Musculoskeletal	21	7.6
Pulmonary	54	19.5
Frontline therapy of cGVHD		
PRD alone	46	16.6
PRD + CNI	99	35.7
PRD + AZA	48	17.3
PRD + MMF	13	4.7
PRD + CNI + AZA	50	18.1
PRD + CNI + MMF	11	4.0
PRD + others	10	3.6

AML indicates acute myelogenous leukemia; ALL, acute lymphoblastic leukemia; AZA, azathioprine; CML, chronic myelogenous leukemia; CLL, chronic lymphocytic leukemia; CNI, calcineurin inhibitors including cyclosporine or tacrolimus; GIT, gastrointestinal tract; MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm; MTX, methotrexate; MMF, mycophenylate mofetil; PRD, prednisone; TCD, T cell depletion.

univariate analysis, we conducted a multivariate analysis using the CIBMTR risk score and GSS risk score as covariates in addition to the cGVHD subtype (classical versus overlap syndrome) and the source of stem cells (peripheral blood versus bone marrow). Given that the CIBMTR risk score is a composite parameter, all the variables incorporated into the CIBMTR risk score were precluded from the multivariate analysis. Multivariate analysis confirmed that severe-grade cGVHD by the GSS and the RG3 to RG6 by the CIBMTR risk score are independently associated with worse FFS, worse OS, and higher NRM in the final multivariate model (Table 3). The group with a high CIBMTR risk score (defined as RG3 to RG6) had a hazard rate (HR) of 1.7 (95% CI, 1.1 to 2.6) for FFS, HR of 4.5 (95% CI, 2.5 to 8.0) for OS, and HR of 4.1 (95% CI, 2.1 to 7.9) for NRM compared with those with a low CIBMTR risk score (defined as RG1 to RG2). The group with severe-grade GSS had an HR of 1.6 for FFS, HR of 2.5 for OS, and HR of 3.3 for NRM compared with those with mild- or moderate-grade GSS (Table 3).

Association of the CIBMTR cGVHD Risk Score and the GSS by the NIH Consensus Criteria

Next, we assessed for any strong linkage between the CIBMTR risk score and the GSS by the NCC. As shown in Table 4, there is a trend of a higher CIBMTR risk score in the patients with more severe-grade GVHD by the GSS, but it was not statistically significant ($P = .075$). Prognostic independence of the CIBMTR risk score and GSS by the NCC suggests that the combination of these 2 scores could potentially yield valuable added prognostic information to either of the individual scores.

To develop and test an improved prognostic system for cGVHD using a combination of the CIBMTR and GSS risk scores, we divided all the patients into 9 subgroups according to their CIBMTR risk score (RG1 versus RG2 versus RG3 to RG6) and GSS (mild versus moderate versus severe grade). We then evaluated the FFS, OS, and NRM in each of the resulting subgroups. We found that in a subgroup analysis confined to the same GSS group, the CIBMTR risk score group (RG) could stratify the patients according to their FFS, OS, or NRM (Figure 3), except in a subgroup with moderate-grade GSS. Specifically, in a subgroup with mild-grade GSS, the RG could stratify the patients

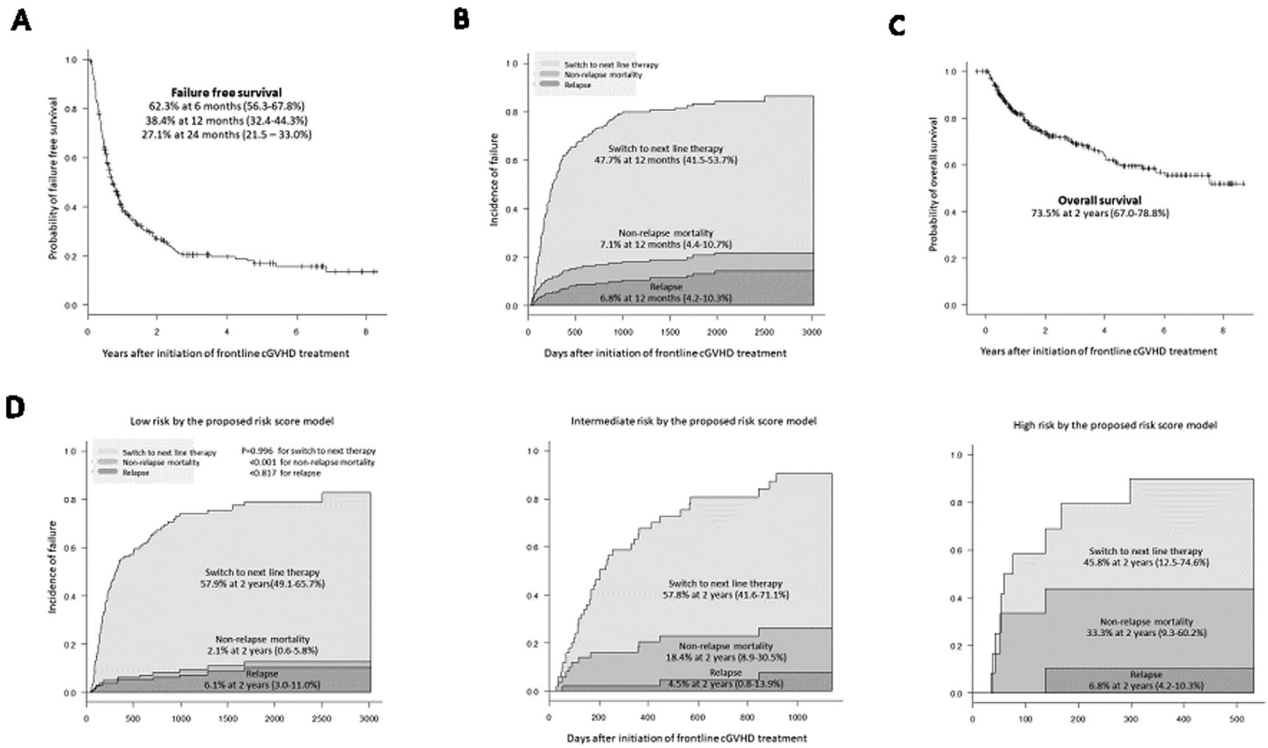

Figure 1. Clinical outcomes as measured in the current patient cohort for failure-free survival (FFS) (A, B) and overall survival (C). The components making up the failure-free survival are shown in panel B, with switch to next-line immunosuppression therapy being the most common cause of treatment failure in the current patient cohort. (D) The combined novel prognostic risk score is most effective at stratifying the FFS with respect to nonrelapse mortality (NRM). When comparing the elements constituting the FFS between the low, intermediate, and high prognostic risk groups within the novel combined risk score, it is the NRM that is statistically different between the risk groups ($P < .003$). No significant difference in the cause of failure among the 3 novel risk groups was noted in the current patient cohort with respect to switch to new/next-line therapy ($P = .996$) or relapse ($P = .817$).

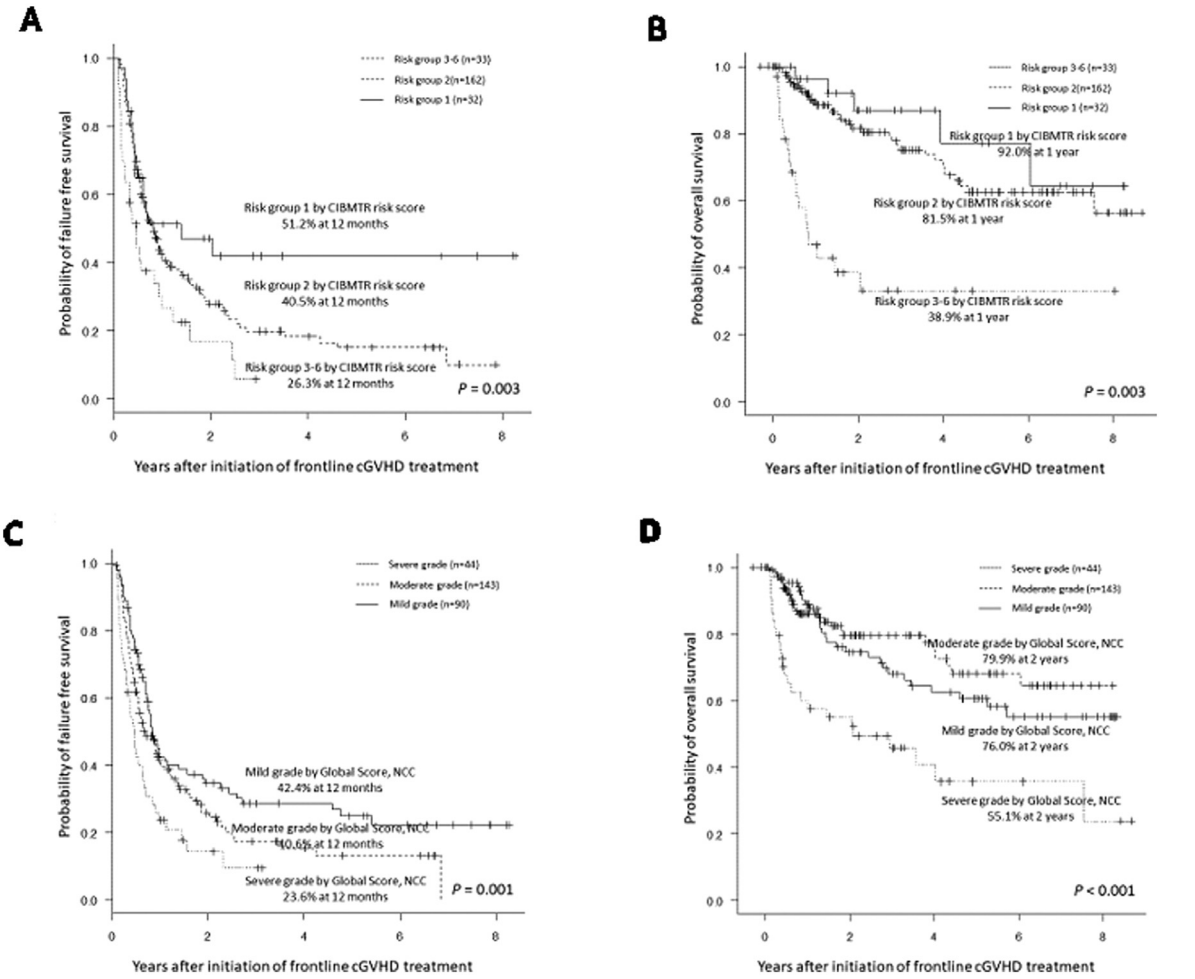

Figure 2. Risk stratification of the patient cohort with respect to clinical outcomes using the individual Centre for International Blood and Marrow Transplant Registry (CIBMTR) and Global Severity Score (GSS) by the National Institutes of Health consensus criteria (NCC) risk scores. The CIBMTR risk score classified most patients in risk group 2 and was able to yield prognostically distinct patient groups with respect to both failure-free survival (A) and overall survival (B). The GSS by NCC risk score was also able to yield statistically significantly different prognostic risk groups with respect to both failure-free survival (C) and overall survival (D).

for FFS ($P = .021$), OS ($P = .008$), or NRM ($P = .017$), whereas in a subgroup with severe-grade GSS, the CIBMTR RG could stratify them for FFS ($P = .036$), OS ($P < .001$), or NRM ($P < .001$).

Proposed Risk Score Model for Outcomes Prediction after Onset of cGVHD

Based on the stratification of the GSS by the CIBMTR RGs, we generated a novel risk score model that combines the GSS and the CIBMTR scores at diagnosis of cGVHD. Specifically, 0 points are given for a CIBMTR risk score of 0 to 6 (RG1 to RG2), whereas 1 point is given for a CIBMTR score of 7 to 12 (RG3 to RG6). Similarly, mild- and moderate-grade cGVHD as assessed with the GSS by the NCC are each given 0 points in the novel risk scoring system, whereas severe-grade cGVHD as assessed with the GSS by the NCC is given 1 point. The final risk score is calculated as a sum of the points assigned by the individual CIBMTR and GSS risk scores as described.

When this scoring system was applied to our patient cohort, we were able to generate the following RGs: score 0 classified as low risk (n = 163), score 1 classified as intermediate risk (n = 50), and score 2 classified as high risk (n = 12). The resulting risk score model could stratify the patients according to their FFS ($P < .001$), OS ($P < .001$), and NRM ($P < .001$). The group with high risk showed significantly worse FFS, worse OS, and higher NRM rate compared with the intermediate and low RGs (Figure 3). As shown in Table 5, the group classified as intermediate risk using the novel score was able to identify patients within CIBMTR RG1 to RG2 who had a lower 6-month and 12-month FFS, lower 2-year OS, and lower 2-year NRM compared with the group classified as low risk ($P < .001$).

The results of the NRI statistic when applied to the addition of the NIH GSS to the CIBMTR risk score in the patient cohort are shown in Table 6. The results suggest that for patients who have attained the clinical event corresponding to the 2-year NRM, the novel score correctly reclassifies 41% of the patients to a higher risk category compared with the CIBMTR at the expense of incorrectly reclassifying 19% of the patients to a lower risk category. This results in an improvement of the novel score compared with the CIBMTR of 22% with respect to the NRM. Similarly, for patients who attained the clinical events corresponding to FFS and OS, the novel score correctly reclassifies 33% and 36% of the patients to a higher risk category compared with the CIBMTR at the expense of incorrectly reclassifying 9% and 23% of the patients to a lower risk category, respectively. Although the new score performs better than the CIBMTR in correctly classifying patients who attain the clinical event of interest, it does not perform as well as the CIBMTR in correctly classifying patients who do not attain the clinical event of interest. This is shown with the negative percentages in the NR(−) column in Table 6.

Table 2
Univariate Analysis for FFS, OS, and NRM after Frontline Treatment for Chronic GVHD

Risk Factor	Category	FFS Duration, Median Days (95% CI)	FFS Rate (% mean ± SD)			OS (%) at 2 yr (95% CI)		NRM (%) at 2 yr (95% CI)	
			6 mo	12 mo	*P* Value		*P* Value		*P* Value
CIBMTR risk score	RG1 (score 0-2; n = 32)	501 (0-1083)	64.9 ± 8.5	51.2 ± 9.1	.003	92.0 (71.5-98.0)	<.001	3.7 (0.3-16.2)	<.001
	RG2 (score 3-6; n = 162)	291 (216-365)	65.2 ± 3.8	40.5 ± 4.1		81.5 (73.3-87.4)		13.3 (8.0-20.0)	
	RG3-6 (score 7-12; n = 33)	166 (94-237)	40.9 ± 8.8	26.3 ± 8.2		38.9 (21.2-56.2)		51.3 (31.0-68.3)	
Chronic GVHD global severity score	Mild grade (n = 90)	304 (237-371)	73.1 ± 4.7	42.4 ± 5.4	.001	74.5 (62.9-83.0)	<.001	10.2 (4.7-18.2)	<.001
	Moderate grade (n = 143)	247 (153-340)	61.6 ± 4.1	40.6 ± 4.3		79.5 (70.1-86.2)		13.7 (7.9-21.1)	
	Severe grade (n = 44)	164 (122-205)	42.5 ± 7.5	23.6 ± 6.5		55.0 (38.7-68.6)		40.3 (25.3-54.9)	
Missing*	n = 50								

* The risk score was calculated with available data for 227 patients. There were 50 patients with a diagnosis of cGVHD for whom the CIBMTR and GSS risk scores could not be calculated retrospectively. These patients were excluded from the analysis shown in Tables 3 and 4 and from the determination of the new cGVHD risk score.

DISCUSSION

Although both the CIBMTR and the GSS by the NCC risk scores have been shown to provide relevant long-term prognostic stratification value in patients with cGVHD, these risk scores have limitations [7,16]. For example, patients within the same GSS prognostic RG can have significantly different long-term outcomes depending on the timing of cGVHD development [17]. This observation can be partially accounted for by the CIBMTR, whereby 0 points are given for cGVHD developing more than 5 months post-transplant as opposed to 1 point being given for cGVHD that develops less than 5 months post-transplant [3,4]. Analogously, the modern-day application of the CIBMTR risk score can yield a substantial proportion of patients within a single RG [7], suggesting a need for further refinement of the risk score.

The present study reports a novel prognostic risk stratification system that combines the GSS by the NCC and CIBMTR risk scores. The potential of a new score developed in this way to add further prognostic information to the individual GSS by the NCC and CIBMTR risk scores is supported by the observation that the aforementioned 2 risk scores are complementary to each other in terms of the parameters they are based on. The CIBMTR risk score is based on factors antecedent to cGVHD development and includes host factors affecting long-term outcomes during GVHD therapy such as performance status [4]. In contrast, the GSS by the NCC is a more precise parameter for biologic cGVHD severity irrespective of antecedent risk factors or host factors [5,9].

The work herein also confirms that both the CIBMTR and the GSS by the NCC risk scores correlate with OS and NRM following first-line systemic treatment for cGVHD with corticosteroid-based regimens, respectively. Furthermore, multivariate analysis suggested that the CIBMTR risk score and the GSS by the NCC independently add prognostic information to patients with cGVHD, and combined use of the 2 scoring systems to generate a new risk score could improve prognostic stratification of patients with cGHVD receiving frontline treatment with respect to OS and NRM. The combination of the GSS and CIBMTR risk scores in our patient cohort was effective in identifying the highest risk patients with the worst OS and NRM. Although this highest RG is of relatively small size in the current cohort, the significantly worse clinical outcomes of these patients necessitate the development of novel therapy and stronger consideration for their involvement in clinical trials. However, this must be done judiciously because the inclusion of patients with such high rates of NRM and treatment failure in early phase clinical trials may result in the early abandonment of otherwise promising new agents. The novel risk score was also able to add prognostic information to those patients who fall into CIBMTR RG 2, providing a needed refinement to the RG into which most patients are classified by the CIBMTR [7].

In addition to adding valuable prognostic information with respect to OS and NRM, the novel combined prognostic risk score developed from a combination of the CIBMTR and GSS by the NCC risk scores was shown to be able to risk stratify patients effectively with respect to FFS. This makes the score amenable to implementation into clinical trial design for the stratification of patients with cGVHD into relevant prognostic groups. Of the factors inherent in the FFS calculation, NRM was best correlated with the combined prognostic risk score (Figure 1). The combined score was better than the CIBMTR or GSS scores in distinguishing those patients with the highest NRM, as the high-risk group in the novel score had a 75% NRM at 2 years.

Table 3
Multivariate Analysis for FFS after Frontline Treatment for cGVHD

Category	FFS		OS		NRM	
	P Value	HR (95% CI)	*P* Value	HR (95% CI)	*P* Value	HR (95% CI)
CIBMTR risk score						
RG1-RG2	.012	1.000	<.001	1.000	<.001	1.000
RG3-RG6		1.719 (1.124-2.628)		4.536 (2.548-8.073)		4.151 (2.165-7.960)
cGVHD global score						
Mild/moderate	.01	1.000	<.001	1.000	<.001	1.000
Severe		1.655 (1.126-2.432)		2.569 (1.509-4.373)		3.324 (1.765-6.259)
cGVHD subtype	.030	1.449 (1.036-2.027)	.939	0.875 (0.507-1.509)	.240	1.529 (0.750-3.116)
Source of stem cells	.054	3.086 (0.980-9.720)	.631	1.057 (0.255-4.380)	.700	1.444 (0.252-9.262)

Table 4
Correlation between the CIBMTR and GSS by NCC Risk Scores (*P* = 0.075*)

CIBMTR cGVHD risk Score	cGVHD global severity score		Severe (n = 41; 18.1%)
	Mild (n = 69; 30.4%)	Moderate (n = 117; 51.5%)	
RG1 (score 0-2; n = 32)	11 (15.9)	18 (15.4)	3 (7.3)
RG2 (score 3-6; n = 162)	50 (72.5)	86 (73.5)	26 (63.4)
RG3-6 (score 7-12; n = 33)	8 (11.6)	13 (11.1)	12 (29.3)**

* By Fisher's exact test.

** The shaded squares correspond to the risk group severity in the novel prognostic risk score. The 12 patients in this darkest shaded square correspond to the high-risk patients in the newly proposed score. The lighter shaded squares correspond to the 165 low-risk patients and the remaining patients in the slightly darker shaded squares correspond to the 50 intermediate-risk patients in the newly proposed score.

A formal comparison of the new risk model with the CIBMTR demonstrates advantages and disadvantages to its implementation in our patient cohort. The novel score performs better than the CIBMTR with respect to predicting the patients with worse 12-month FFS, 2-year NRM, and 2-year OS. This implies that for the patients in the current cohort who experienced the clinical outcome events of treatment failure, NRM, and death, the novel score was more likely to classify them correctly in a higher risk category than the CIBMTR. However, for the patients in the current cohort who did not attain the clinical outcomes, the novel score was less likely than the CIBMTR to classify them correctly into a lower risk category.

The novel score is therefore better able than the CIBMTR to identify patients with poorer clinical outcomes. This means that it is likely to upstage more appropriately a subset of the patients in CIBMTR RG2 who have worse clinical outcomes than would be predicted with the CIBMTR. However, as is shown with the net reclassification improvement test, this comes at the expense of having a risk of incorrectly classifying some of the CIBMTR RG2 patients as having poorer expected clinical outcomes. Although the net risk of incorrect reclassification is low (6% for 2-year NRM and 8% for 2-year OS as shown in Table 6), this must be kept in mind when selecting a prognostic risk score to apply to patients with cGVHD who start on systemic therapy.

Some important limitations to the net reclassification improvement must be considered. First, the NRI does not comment on the efficacy of the risk model but is intended to evaluate the classification improvement of the addition of more data (the NIH GSS) to the existing risk model (CIBMTR). However, the NRI does not provide any information on how well the new risk score is calibrated to the outcomes in the patient cohort. Because the CIBMTR was initially derived in a patient cohort different from the new risk score, the risk of calibration error is significant. In our study, this risk is mitigated by comparing the new risk score with the CIBMTR when applied to the same patient cohort. Furthermore, the cutoff of evaluating

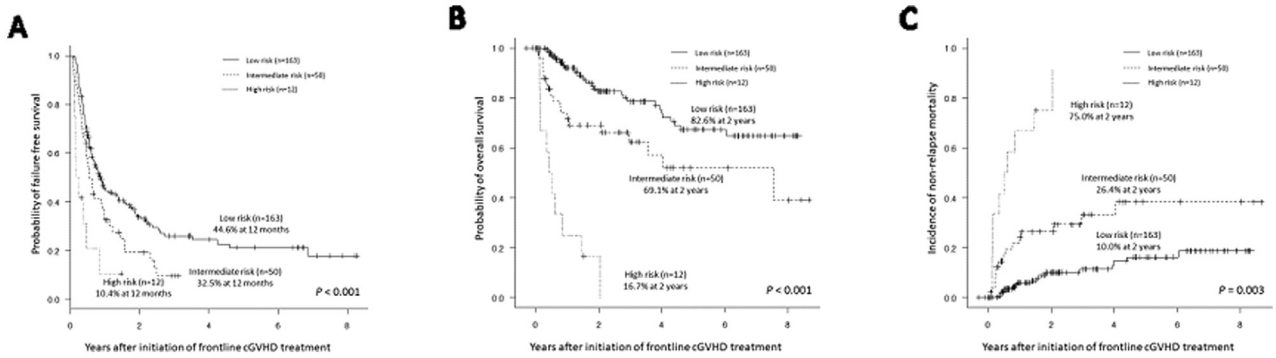

Figure 3. Novel prognostic risk score incorporating the Centre for International Blood and Marrow Transplant Registry and Global Severity Score by the National Institutes of Health consensus criteria prognostic risk groups. (A) The resulting high-risk group had a 12-month failure-free survival of 10.4% as opposed to 32.5% in the resulting intermediate-risk group and 44.6% in the resulting low-risk group. (B) The high-risk group had a 2-year overall survival of 16.7%, significantly less than 69.1% in the intermediate-risk group and 82.6% in the low-risk group. (C) Similarly, the high-risk group had a 2-year nonrelapse mortality of 75.0%, significantly higher than 26.4% in the intermediate-risk group and 10.0% in the low-risk group.

Table 5
Proposed Risk Score Model for FFS, OS, and NRM after the Frontline Treatment for cGVHD

				FFS			OS	NRM
Risk Group	CIBMTR cGVHD Risk Score	cGVHD GSS	No. of Patients	Median	6 mo	12 mo	2 yr	2 yr
Low risk	RG1-RG2	Mild-moderate grade	165	311 (230-500)	67.0 (59.2-73.6)	44.6 (36.5-52.4)	82.6 (74.1-88.5)	10.0 (5.4-16.3)
Intermediate	RG1-RG2, RG3-RG6	Severe grade, mild-moderate grade	50	203 (150-357)	55.7 (40.9-68.2)	32.5 (19.9-45.8)	69.1 (53.3-80.5)	26.4 (14.4-39.9)
High risk	RG3-RG6	Severe grade	12	59 (38-166)	20.8 (3.5-47.9)	10.4 (0.6-36.4)	16.7 (2.7-41.3)	75.0 (35.2-92.4)
P value				<.001			<.001	<.001

the new risk score as having an improvement of at least 5% over the CIBMTR in correctly classifying patients does not have a published clinical significance but was selected as a cutoff of comparing the risk models. Additional comparisons between the 2 scores should be performed once the novel score is validated in a larger cohort of patients.

Several important limitations of the novel prognostic risk score model must be taken into account. First, the risk score was developed in a retrospective cohort of patients with cGVHD during a period when diagnosis of cGVHD was based on the 2005 NIH consensus criteria [2]. The criteria were further refined in 2014 and have resulted in some changes to the determination of cGVHD severity based on individual organ systems [18]. However, because the cGVHD criteria changed relatively little between 2005 and the 2014 update [9], the more recent refinement to the 2005 NIH criteria is unlikely to alter significantly the prognostic efficacy of the new prognostic score in this patient cohort. This is also because most cGVHD severity in the cohort was caused by skin and liver cGVHD manifestations, with relatively fewer patients having lung manifestations. For those patients who did have lung manifestations, the scoring was largely based on the Forced Expiratory Volume in 1 second, which yields similar risk categories in both the 2005 version of the score and the 2014 update. For the patients in this cohort who had liver manifestations of cGVHD, the 2014 update would change a total of 9% of all cumulative cGVHD scores as a result of a change in the liver score. Within the CIBMTR RG3 to RG6, this would result in a change from moderate cGVHD to mild cGVHD in 9% of the cases, whereas within the CIBMTR RG2, it would result in a change from moderate to mild cGVHD in 12% of cases. Within the CIBMTR RG1, 3% of cases change from moderate to mild cGVHD and 7% of cases change from mild to moderate cGVHD, resulting in a total change of 10% of cases in this RG. The relatively small change in cGVHD scores is equally distributed among the CIBMTR groups and would therefore be expected to have the same relative effect on each of the groups formed by the novel score (being a composite of the CIBMTR and NIH GSS scores). We expect, therefore, that the overall risk stratification of the novel score with respect to FFS, NRM, and OS would not be changed significantly by the 2014 update to the NIH GSS score.

Another limitation of the current study refers to the greater risk of bias inherent in a single-center retrospective cohort that may have affected diagnostic, monitoring, and treatment decisions in terms of switching cGVHD therapy that are important for the calculation of FFS. The single-center derivation limits the prognostic predictive capacity of the score for any single patient. In addition, the retrospective nature of this study may result in clinical outcomes within prognostic risk score groups that differ from the outcomes in corresponding groups in larger prospective cohorts. For example, the patients in this cohort with a mild GSS score appeared to have a slightly shorter 2-year OS and NRM than the patients with moderate GSS. This argues for further validation of the novel combined risk score in a larger prospective multicenter cGVHD patient cohort before its implementation in future clinical trials.

CONCLUSION

The CIBMTR risk score and the GSS by the NCC are complementary but well correlated with FFS, OS, and NRM following frontline systemic treatment for cGVHD with corticosteroid-based regimens. A new risk score model combining the CIBMTR risk score and the GSS by the NCC has been developed and may enhance prognostic stratification in patients with cGVHD receiving frontline cGVHD treatment.

Table 6
Category-Based Net Reclassification Index Applied to the New Risk Model in Comparison with the CIBMTR Risk Score Model

	NRI (+), %*	Pr (up|event)†	Pr(down|event)	NRI (−), %	Pr(up|nonevent)	Pr(down|nonevent)‡
12-month FFS	+24	0.33 (0.13-0.41)	0.09 (0.001-0.7)	−18	0.26 (0.09-0.36)	0.08 (0.001-0.77)
2-year NRM	+22	0.41 (0.24-0.54)	0.19 (0.05-0.35)	−6	0.13 (0.08-0.18)	0.068 (0.03-0.11)
2-year OS	+13	0.36 (0.14-0.49)	0.23 (0.10-0.39)	−8	0.13 (0.01-0.29)	0.05 (0.02-0.09)

* NRI(+) refers to the net percentage of patients with the event of interest correctly reclassified to a higher risk by the new model. Similarly, NRI(−) refers to the net percentage of patients without the clinical event of interest correctly reclassified to a lower risk by the new model. A negative percentage implies that a greater number of patients was reclassified incorrectly than correctly.

† Pr(up|event) refers to the probability of a patient being correctly reclassified up to a higher risk by the new model given the event occurred. The event here refers to the relevant clinical outcome, corresponding to the 12-month FFS, 2-year NRM, and 2-year OS. Note that Pr(down|event) refers to the probability that a patient was incorrectly reclassified down to a lower risk by the new model given the event occurred. Probabilities are reported with 95% confidence intervals.

‡ Pr(down|nonevent) refers to the probability of a patient being correctly reclassified down to a lower risk by the new model given the event occurred. Pr(up|nonevent) refers to the probability of a patient being incorrectly reclassified up to a higher risk by the new model given the event did not occur.

ACKNOWLEDGMENTS

Financial disclosure: The authors have nothing to disclose.

Conflict of interest statement: There are no conflicts of interest to report.

REFERENCES

1. Arai S, Arora M, Wang T, et al. Increasing incidence of chronic graft-versus-host disease in allogeneic transplantation—a report from CIBMTR. *Biol Blood Marrow Transplant*. 2015;21(2):266–274.
2. 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;11:945–955.
3. Arora M, Klein JP, Weisdorf DJ, et al. Chronic GVHD risk score: a Centre for International Blood and Marrow Transplant Research analysis. *Blood*. 2011;117(24):6714–6720.
4. Arora M, Hemmer MT, Ahn KW, et al. CIBMTR chronic GVHD risk score predicts mortality in an independent validation cohort. *Biol Blood Marrow Transplant*. 2015;21(4):640–645.
5. Pidala J, Kim J, Anasetti C, et al. The global severity of chronic graft-versus-host disease, determined by National Institutes of Health consensus criteria, is associated with overall survival and non-relapse mortality. *Haematologica*. 2011;96(11):1678–1684.
6. Moon JH, Hamad N, Sohn SK, et al. Improved prognostic stratification power of CIBMTR risk score with the addition of absolute lymphocyte and eosinophil counts at the onset of chronic GVHD. *Ann Hematol*. 2017;96:805–815.
7. Inamoto Y, Kim D, Storer BE, et al. Application of CIBMTR risk score to NIH chronic GVHD at individual centres. *Blood*. 2014;123:453–455.
8. Aisa Y, Mori T, Kato J, et al. Validation of NIH consensus criteria for diagnosis and severity-grading of chronic graft-versus-host disease. *Int J Hematol*. 2013;97:263–271.
9. Lee SJ. Classification systems for chronic graft-versus-host disease. *Blood*. 2017;129(1):30–37.
10. Arai S, Jagasia M, Storer B, et al. Global and organ-specific graft-versus-host disease severity according to the 2005 NIH consensus criteria. *Blood*. 2011;118(15):4242–4249.
11. Inamoto Y, Storer BE, Lee SJ, et al. Failure-free survival after second-line systemic treatment of chronic graft-versus-host disease. *Blood*. 2013;121(12):2340–2346.
12. Inamoto Y, Flowers MED, Sandmaier BM, et al. Failure-free survival after initial systemic treatment of chronic graft-versus-host disease. *Blood*. 2014;124(8):1363–1371.
13. Palmer J, Chai X, Pidala J, et al. Predictors of survivals, nonrelapse mortality, and failure-free survival in patients treated for chronic graft-versus-host disease. *Blood*. 2016;127(1):160–166.
14. Leening MGJ, Vedder MM, Witteman JCM, et al. Net reclassification improvement: computation, implementation, and controversies. *Ann Intern Med*. 2014;160:122–131.
15. Kanda Y. Investigation of the freely available easy-to-use software "EZR" for medical statistics. *Bone Marrow Transplant*. 2013;48:452–458.
16. 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;13:1207–1215.
17. 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;14:1163–1171.
18. Martin PJ, Lee SJ, Przepiorka D, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease, VI: the 2014 clinical trial design working group report. *Biol Blood Marrow Transplant*. 2015;21:1343–1359.