Biol Blood Marrow Transplant 25 (2019) 2468–2473

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

Survivorship

The Chronic Graft-versus-Host Disease Failure-Free Survival (cGVHD-FFS) Index

Joseph A. Pidala[1], Betty K. Hamilton[2], Paul J. Martin[3,4], Lynn Onstad[3], Barry E. Storer[3], Jeanne Palmer[5], Amin Alousi[6], Corey Cutler[7], Madan H. Jagasia[8], George L. Chen[9], Mukta Arora[10], Mary E. Flowers[3,4], Stephanie J. Lee[3,4,*]

[1] *H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[2] *Blood and Marrow Transplantation, Department of Hematology and Medical Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio*
[3] *Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, Washington*
[4] *Department of Medicine, University of Washington, Seattle, Washington*
[5] *Mayo Clinic, Phoenix, Arizona*
[6] *Department of Stem Cell Transplant and Cellular Therapy, University of Texas, MD Anderson Cancer Center, Houston, Texas*
[7] *Division of Hematologic Malignancies, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts*
[8] *Division of Hematology/Oncology, Vanderbilt-Ingram Cancer Center, Nashville, Tennessee*
[9] *Department of Medicine, Roswell Park Cancer Institute, Buffalo, New York*
[10] *Hematology, Oncology and Transplantation, University of Minnesota, Minneapolis, Minnesota*

Article history:
Received 3 June 2019
Accepted 25 July 2019

Keywords:
Chronic graft-versus-host disease
Failure-free survival
Prognostic model

A B S T R A C T

In clinical trials of chronic graft-versus-host disease (cGVHD), the need to start a new systemic treatment is considered a treatment failure. A composite endpoint called "failure-free survival" (FFS), where events are initiation of a new systemic cGVHD treatment, recurrent malignancy, and death, has been suggested as a possible long-term indicator of success. The goal of the current study was to identify changes in cGVHD manifestations from baseline to 6 months that could accurately predict subsequent longer-term FFS, thereby making it possible to assess outcomes earlier than would otherwise be possible. We used data from 2 prospective, multicenter, observational studies to develop the cGVHD-FFS index. The cGVHD-FFS index was calculated at 6 months, a typical time-point for assessment of the primary endpoint of phase II cGVHD trials. Subsequent FFS was only 45% within the next 2 years. We found that changes in the scores for the eyes, joint/fascia, and mouth ulcers from baseline to 6 months were associated with subsequent FFS, but the prognostic accuracy of these changes was not adequate for use in trials. Biomarker studies might help to identify criteria that improve prediction of long-term clinical outcomes in patients with cGVHD.

INTRODUCTION

Chronic graft-versus-host disease (cGVHD) is the most serious and common long-term complication of allogeneic hematopoietic cell transplantation, occurring in 20% to 50% of people [1,2]. One of the critical barriers in cGVHD research is the lack of validated intermediate endpoints, meaning that clinical trials require long follow-up to observe endpoints of interest [3]. As a consequence, clinical trials become prohibitively long and expensive. To address this gap in the field, we sought to develop a new tool called the cGVHD failure-free survival index (cGVHD-FFS) that will serve as an intermediate endpoint to predict subsequent failure-free survival (FFS), defined as absence of death, relapse, and requirement for new systemic cGVHD therapy (i.e., events that constitute treatment failures in clinical trials).

Most of the failure events in the cGVHD population represent change of therapy, which is a poor prognostic sign [4,5]. The goal of the current study was to identify changes in cGVHD manifestations from baseline to 6 months that could accurately predict subsequent longer-term FFS, thereby making it possible to assess outcomes earlier than would otherwise be possible. We anticipated that the results would also help identify the organs most associated with subsequent treatment failure.

Financial disclosure: See Acknowledgments on page 2472.

* Correspondence and reprint requests: Stephanie J. Lee, MD, MPH, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, D5-290, Seattle, WA 98109.
E-mail address: sjlee@fredhutch.org (S.J. Lee).

MATERIALS AND METHODS

Cohort

Data from 312 patients were derived from 2 prospective, longitudinal, observational cohorts (NCT00637689 and NCT01902576). Both studies included patients with prior allogeneic stem cell transplant from any graft

https://doi.org/10.1016/j.bbmt.2019.07.040
1083-8791/

source, donor type, and graft-versus-host disease (GVHD) prophylaxis. Patients had to meet diagnostic definitions according to the cGVHD National Institutes of Health (NIH) consensus criteria [6,7] and either started or planned to start a new systemic immunosuppressive agent, called the "index treatment," within 4 weeks (required for enrollment in 1 study; a subset selected from the other study). Thus, classic and overlap subtypes of cGVHD were included, whereas purely late acute GVHD was not. Systemic treatment was defined as any medication or intervention with intended systemic effects, including extracorporeal photopheresis. The 2 prior cohorts included both first-line cGVHD therapy, as well as subsequent lines of therapy (Table 2). Topical therapies and fluticasone, azithromycin, and montelukast were not considered systemic treatments. The study was approved by the institutional review boards of participating centers, and all participants provided signed informed consent.

Data Collection

Data used in this analysis were collected from laboratory testing and reported by providers and patients at enrollment and 6 months later. Charts were reviewed annually for relapse, death, addition of a new immunosuppressive treatment, or discontinuation of immunosuppression after a patient completed the primary study.

Laboratory Testing

Two laboratory measures (liver function testing and the percent predicted forced expiratory volume in the first second from pulmonary function tests) were recorded.

Clinician Assessments

Candidate variables include the provider and laboratory measures recommended for collection by the NIH cGVHD consensus conferences [6-9]. These included the 8 provider-reported NIH organ severity scores (skin, eye, mouth, genital, gastrointestinal, lung, liver, and joint). For the earlier cohort, the skin score was the higher of the body surface area and sclerotic scores. The NIH response items that differ from the scoring variables (mouth, esophagus, upper and lower gastrointestinal [GI] tract, joint) were also included as potential explanatory variables.

Patient Assessments

Patients reported their cGVHD symptoms using the Lee symptom scale (30 items, captured as a 5-point Likert scale from *no symptoms* to *extremely bothered*). The summary score was calculated according to the recommendations of the developer [10].

Chart Review

Chart review captured FFS, defined as the time in days from enrollment until death, relapse of the underlying malignancy, or the addition of a new systemic cGVHD treatment. Addition of a new systemic immunosuppressive therapy was defined as a new medication or intervention to treat cGVHD added 4 or more weeks after enrollment (because 1 study allowed enrollment up to 4 weeks before adding a new medication). However, if cGVHD progression was the reason for adding a new agent within 4 weeks of enrollment, the event was considered a failure. Switching medications due to toxicity, insurance approval or financial considerations, or procedural complications (eg, loss of intravenous access for extracorporeal photopheresis) without a corresponding worsening of cGVHD was not considered a failure. Addition of an agent for "steroid sparing" more than 4 weeks after the start of the index treatment was considered a failure. However, only 6 of 88 (7%) new medications were reported as being started for steroid sparing among the patients for whom we collected the reason for starting a new treatment. FFS was censored at last follow-up or if patients were lost to follow-up.

Biostatistical Analysis

The analysis was limited to patients who had both enrollment and 6-month assessments and who had not already failed the index treatment by 6 months because of relapse or progressive malignancy or having started a new systemic treatment. Patients who had an event before 6 months were excluded because they would already be considered treatment failures in a clinical trial with a 6-month endpoint; our goal was to predict FFS after 6 months. For purposes of this analysis, we assumed that a new systemic treatment started within 2 weeks after the 6-month assessment was prompted by inadequate response at 6 months, and these patients were also excluded.

Cox regression analysis of FFS subsequent to the 6-month assessment point was used to identify prognostic variables. Stepwise regression was used to identify the most prognostic variables, and an index was created based on the log hazard ratios for the included variables. For purposes of evaluating the suitability of the cGVHD-FFS index as a surrogate endpoint for clinical trials, the index defined in the final model was evaluated in a logistic regression model for FFS as a binary outcome 6 months after the 6-month assessment (ie, 12 months after enrollment), excluding censored observations. The c-statistic (area under the receiver operating characteristic curve) was used to quantify the prognostic ability.

In secondary analyses, we tested whether the 6-month cGVHD-FFS index predicted subsequent nonrelapse mortality or relapse and whether the prognostic power of the cGVHD-FFS index differed in defined patient subgroups such as initial versus subsequent therapy, prednisone-based index treatment, and cGVHD severity at enrollment.

RESULTS

Cohort Characteristics

Table 1 and Figure 1 show the cohort characteristics. Patients (n = 312) in this analysis were enrolled at 14 institutions in the United States and were evenly split between related and unrelated donors, myeloablative and less intensive conditioning regimens, and myeloid malignancies versus other types of diseases. Most (93%) received mobilized peripheral blood.

Table 2 shows the cGVHD characteristics. Approximately 60% were enrolled within 3 months of diagnosis (63% incident cases)

Table 1
Demographic Characteristics of Patients (n = 312)

Characteristic	Value
Age at initial treatment, median (range), yr	54 (2-79)
Patient sex, n (%)	
Male	193 (62)
Female	119 (38)
Donor-patient sex combination, n (%)	
Female to male	79 (26)
Other	230 (74)
White, n (%)	280 (90)
Hispanic, n (%)	17 (6)
Diagnosis, n (%)	
Myeloid malignancy	172 (55)
Lymphoid malignancy	105 (34)
Other/nonmalignant	35 (11)
Conditioning regimen, n (%)	
High dose with or without TBI	148 (48)
Reduced intensity or nonmyeloablative	163 (52)
Graft source, n (%)	
Bone marrow	15 (5)
Mobilized blood cells	289 (93)
Cord blood	8 (3)
Donor and HLA type, n (%)	
HLA-matched related	118 (38)
HLA-matched unrelated	143 (46)
HLA antigen or allele-mismatched related	8 (3)
HLA antigen or allele-mismatched unrelated	42 (14)
Transplant center, n (%)	
Fred Hutchinson	144 (46)
Vanderbilt	34 (11)
University of Minnesota	29 (9)
Dana-Farber	26 (8)
Stanford	19 (6)
Cleveland Clinic	14 (4)
University of British Columbia	14 (4)
Moffitt Cancer Center	12 (4)
Roswell Cancer Center	9 (3)
MD Anderson	6 (2)
Medical College of Wisconsin	3 (1)
Other*	2 (1)

TBI indicates total body irradiation.
* Duke, Northwestern Children's Hospital: 1 each.

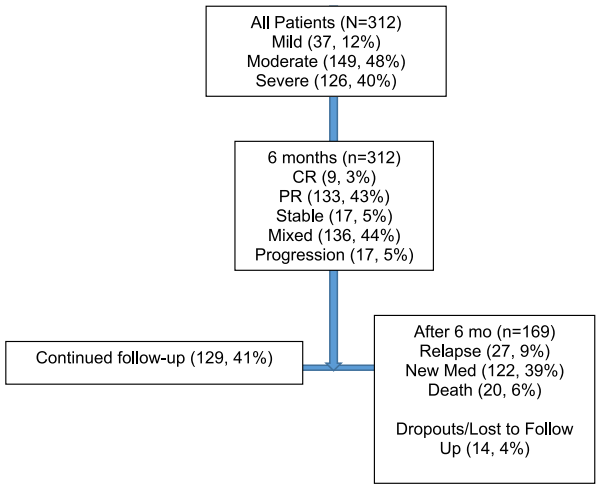

Figure 1. Disposition of the study population.

Table 2
Chronic GVHD Characteristics of Participants (N = 312)

Characteristic	Value
Type of case, n (%)	
Incident (enrolled within 3 months of diagnosis)	198 (63)
Prevalent (enrolled more than 3 months after diagnosis)	114 (37)
Prior grade II to IV acute GVHD, n (%)	134 (44)
NIH global severity, n (%)	
Mild or less than mild	37 (12)
Moderate	149 (48)
Severe	126 (40)
Overlap subtype, n (%)	129 (41)
Organs involved, n (%)	
Skin	215 (69)
Fascia/joints	132 (42)
Mouth	184 (59)
Eye	170 (55)
Lung	78 (25)
GI tract	99 (32)
Liver	63 (21)
Genital	29 (13)
Line of treatment, n (%)	
Initial	189 (61)
Second line	52 (17)
Third line	37 (12)
Fourth line	16 (5)
Fifth line or beyond	18 (6)
Index treatment (started ± 4 weeks of enrollment), n (%)	
Prednisone-based ± others	195 (63)
Mycophenolate mofetil	5 (2)
Sirolimus	17 (5)
Extracorporeal photopheresis	35 (11)
B cell antibody therapy (rituximab, ofatumumab)	13 (4)
Other	47 (15)
Outcomes	
Follow-up for survivors, median (range), mo	29 (3-122)
FFS at 1 year after the 6-month assessment, %	65
FFS at 2 years after the 6-month assessment, %	47
Survival at 1 year after the 6-month assessment, %	95
Survival at 2 years after the 6-month assessment, %	88

and were on initial treatment (61%) and prednisone-based regimens (63%). Extracorporeal photopheresis with or without other agents was the next most common treatment (11%), with sirolimus (5%) and B cell targeted therapy (4%) the next most common regimens. Mild cGVHD was uncommon (12%), and most had moderate (48%) or severe (40%) disease. Skin, mouth, and eye involvement was seen in more than half of participants; other organs were also frequently involved. Median follow-up after the 6-month assessment was 29 months (range, 3 to 122 months).

Univariable Predictors of FFS

Overall and FFS after the 6-month assessment were 95% and 65%, respectively, at 1 year (18 months after enrollment) and 88% and 47% at 2 years (Table 2). We tested whether restricting to a shorter time frame, such as 6 to 12 months after the 6-month assessment and censoring thereafter, would improve the model's prognostic ability. Because the results were not affected, the entire duration of follow-up was used to identify variables associated with subsequent FFS. Table 3 shows the variables whose change, whether improvement or worsening, was significantly associated with subsequent FFS in univariate analyses. Change in oral ulcers, range of motion at the wrist and all joints, forced expiratory volume in the first second, and scores of the eye, joints/fascia, and skin were significantly correlated with subsequent FFS. Notably, change in GI tract measures (liver, esophageal, upper and lower tracts), clinical lung symptoms, mouth summary score, and patient-reported overall symptom burden were not prognostic of subsequent FFS.

We also explored the prognostic ability of a patient's status at 6 months irrespective of change since baseline. Results are shown in the last 3 columns of Table 3. Here, the mouth variables (erythema, lichen-planus-like changes, and ulcers) measured at 6 months were statistically significant predictors.

Development of the cGVHD-FFS Index from the Multivariable Model

The best model identified in multivariable analysis was based on changes in the eye, joint, and mouth ulcer scores predicting subsequent FFS (Table 4). Figure 2 shows the FFS curves based on cGVHD-FFS index >0 (n = 78, shorter FFS) versus cGVHD-FFS index ≤ 0 (n = 231, longer FFS). The curve continued to fall linearly, showing that failure was an ongoing problem even after 6 months on the same treatment. There was no apparent flattening or plateau on the curves as one would expect to see if the failure rate was decreasing. The c-statistic for predicting FFS at 6 months after assessment was 0.62 after deleting observations censored before 6 months, illustrating that although the model worked reasonably well for the entire population, it did not accurately predict the outcome for individuals. Limiting the analysis to patients receiving initial prednisone-based therapy, the group in which the index worked best, the c-statistic was still 0.64. A c-statistic of 0.50 means the index is no better than chance at predicting FFS at 1 year after enrollment.

Correlation of the cGVHD-FFS Index with Other Outcomes

The cGVHD-FFS index was statistically associated with failure due to an added medication (n = 120 events) (hazard ratio [HR], 1.91; 95% confidence interval [CI], 1.30 to 2.79; $P = .0009$) but not failure due to relapse (n = 27) (HR, 1.27; 95% CI, 0.50 to 3.21; $P = .61$) or death (n = 20 events) (HR, 2.11; 95% CI, 0.73 to 6.13; $P = .17$), although the small numbers of relapse and death events limited statistical power (Table 5). Accordingly, the

Table 3
cGVHD-FFS Index: Univariate Analysis (169 Events in 312 Patients)

	Change* from Baseline			6-Month Values		
Measure (Scale Range)	Number Evaluated	χ^2	*P*	Number Evaluated	χ^2	*P*
Alkaline phosphatase, U/dL	296	0.07	.79	304	1.42	.23
Alkaline phosphatase/upper limit of normal ratio	296	0.00	.97	304	1.82	.18
Alanine aminotransferase, U/dL	297	0.02	.88	303	0.52	.47
Alanine aminotransferase/ULN ratio	297	0.01	.94	303	1.18	.28
Total serum bilirubin, mg/dL	298	0.29	.59	304	0.04	.85
Total serum bilirubin/ULN ratio	298	0.19	.67	304	0.09	.76
Esophagus score (0-3)	312	0.20	.65	312	2.58	.11
Lower GI score (0-3)	310	3.45	.06	312	0.00	.97
Oral erythema score (0-3)	311	0.25	.61	311	5.91	**.02**
Oral lichenoid score (0-3)	311	0.00	.98	311	5.27	**.02**
Oral sum score (0-12)	312	3.49	.06	312	9.65	**.002**
Oral ulcer score (0-6)	**311**	**5.05**	**.02**	312	5.68	**.02**
Upper GI score (0-3)	312	0.01	.91	312	0.54	.46
Ankle range of motion score (1-4)	262	2.07	.15	278	1.85	.17
Elbow range of motion score (1-7)	267	0.34	.56	280	0.00	1.00
Shoulder range of motion score (1-7)	270	1.28	.26	282	0.08	.78
Wrist and finger range of motion score (1-7)	**258**	**4.26**	**.04**	274	0.67	.41
Joint range of motion sum score (0-25)	**267**	**3.91**	**.05**	280	1.11	.29
Eye score (0-3)	**311**	**8.14**	**.004**	312	1.91	.17
Genital score (0-3)	163	0.03	.86	193	2.15	.14
GI tract score (0-3)	311	0.86	.35	312	0.80	.37
Joints/fascia score (0-3)	**311**	**7.18**	**.007**	311	0.69	.41
Lung score (0-3)	312	0.00	.95	312	0.13	.72
Mouth score (0-3)	312	0.53	.47	312	3.74	**.05**
Skin score (0-3)	**311**	**4.51**	**.03**	311	3.36	.07
Forced expiratory volume in first second, percent of predicted	**97**	**4.70**	**.03**	134	0.11	.75
Summary symptom score reported by patient (0-100)	238	2.49	.11	269	1.24	.27

ULN indicates upper limit of normal.
Numbers in bold indicate statistically significant predictors at p<=0.05.
* Improvement or worsening.

cGVHD-FFS index did not show a statistically significant association with survival or nonrelapse mortality.

We tested the prognostic ability of the cGVHD-FFS index in patient subsets (Table 5) The cGVHD-FFS index seemed most useful in newly diagnosed patients treated with prednisone, but the improved prognostic ability over other patient groups was not statistically significant (interaction $P > .05$). Performance of the cGVHD-FFS index did not differ between patients with mild/moderate cGVHD and those with severe cGVHD.

DISCUSSION

Although we identified several NIH response variables whose change by 6 months was associated with subsequent FFS, the cGVHD-FFS index lacked accuracy and would not be appropriate as an intermediate endpoint, even in early phase clinical trials. With a c-statistic of 0.62, the model is only slightly better than chance when predicting subsequent new systemic therapy, the major cause of failure in our population.

The inability to generate a more prognostic model is disappointing but not surprising given the heterogeneity of cGVHD manifestations. The selected variables were statistically significant for the population but do not adequately predict the outcomes for individuals. Thus, clinical trials interested in FFS

Table 4
Weights per Unit Change in Measures Associated with FFS

Measure*	Weight per Unit Change†
Eye score	0.2814
Joint/fascia score	0.3698
Mouth ulcer score	0.1341

* In each of these scales, higher values reflect more severe manifestations of cGVHD.
† Weights are multiplied by the difference between the 6-month assessment value and the baseline value. For example, the value of a change from a score of 1 at baseline to 3 at 6 months is 2. For the eye score, 2 is multiplied by 0.2814, yielding a weighted score of 0.5628, which is greater than 0 and predictive of a shorter subsequent FFS. The change index is the sum of weighted scores in all 3 measures.

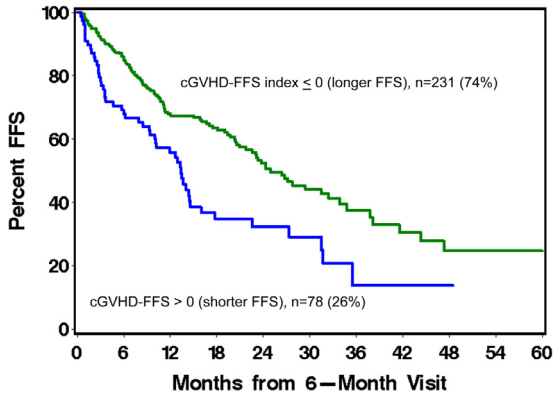

Figure 2. Failure-free survival curves for cGVHD-FFS index >0 (n = 78) versus FFS index ≤0 (n = 231).

Table 5
Summary of Outcomes after Month 6 Visit, Based on cGVHD-FFS Index >0 versus ≤0

Characteristic	No.	No. of Events	HR (95% CI)	*P*	$P_{interaction}$
FFS (all events)	309	167	1.82 (1.30-2.54)	.0005	
New medication		120	1.91 (1.30-2.79)	.0009	
Relapse		27	1.27 (0.50-3.21)	.61	
Death		20	2.11 (0.73-6.13)	.17	
Overall survival	309	56	1.03 (0.56-1.89)	.92	
Nonrelapse mortality	309	40	1.14 (0.57-2.29)	.71	
FFS by line of therapy					
Initial	188	110	2.12 (1.41-3.21)	.0003	.22
Subsequent	121	57	1.36 (0.75-2.45)	.31	
FFS by cGVHD severity					
Mild/moderate	184	103	1.70 (1.13-2.56)	.01	.42
Severe	125	64	2.12 (1.16-3.87)	.01	
FFS by index medication					
Others	115	59	1.46 (0.80-2.67)	.22	.46
Prednisone based	194	108	1.94 (1.29-2.92)	.002	

beyond 6 months as an endpoint will still have to follow the study population to observe these events, or a new model that incorporates additional information could be constructed. Our model worked best for patients receiving initial therapy and prednisone-based treatments, perhaps because population characteristics and management decisions are more uniform in those settings than with longer duration of cGVHD. We acknowledge that the original cohorts that informed this current analysis had heterogeneity in cGVHD features and lines and type of therapies delivered at baseline.

Our results suggest that patients who have worsening in their eyes, joints/fascia, or oral ulcers when assessed at 6 months are more likely to experience subsequent treatment failure. Conversely, we did not find that changes in the skin, liver, lungs, or GI tract at 6 months were prognostic with regard to subsequent treatment failure. Several prior analyses sought correlations between NIH-recommended variables and subsequent outcomes such as complete and partial responses or FFS. One study reported that changes in the eyes, joints, range of motion, lung symptom scores, and lower GI scores were associated with attainment of an NIH-defined complete or partial response [5], and complete and partial responses were associated with a higher likelihood of successful discontinuation of immunosuppression and better survival. Another analysis showed that enrollment scores of skin, GI tract, range of motion, forced vital capacity, liver score, and bronchiolitis obliterans syndrome predicted FFS [11], whereas changes in clinician-assessed skin scores and patient-reported itching at 6 months predicted subsequent FFS after 6 months [12]. In the present analysis, changes in the eyes, joints, and mouth ulcers were prognostic of subsequent FFS, which may not be surprising as these manifestations may be quite symptomatic. These organ variables differ somewhat from the variables identified in the 2 prior studies, but the apparent discrepancies may be explained by the different endpoints and whether only enrollment or 6-month variables were included. Complete and partial responses reflect improvement and are probably managed by medication tapers and discontinuation, whereas cGVHD worsening is treated overwhelmingly with addition of a new systemic medication and is considered a failure event in our analysis. Our clock started at 6 months because we were interested in predicting FFS after that point, whereas other analyses started the clock at enrollment and used only baseline variables.

We assumed that most phase II studies would include at least 6 months of follow-up to assess for an efficacy signal. Our results suggest that FFS beyond 6 months must be observed and cannot be extrapolated from a patient's status at 6 months. We acknowledge that clinician behavior directly affects FFS, because tapering immunosuppression may prompt GVHD worsening and providers have differing thresholds for starting new treatments. In addition, reasons for starting new medications are diverse and may include attempts to "spare steroids," defined as using nonsteroid immunosuppressive agents to try to achieve a lower dose of steroids, although this was an uncommon reason in our study. Providers may start a new agent even if cGVHD manifestations are not worse and they are trying to improve symptoms. Availability of clinical trials or newly available therapies may have also played a role.

Becuase we were not successful generating a prognostic score for subsequent FFS despite a large sample size of 312 patients, we do not believe that additional attempts based solely on clinical factors will be fruitful. We suggest that biologic markers may better identify patients with cGVHD that will continue to be adequately controlled with current treatment.

ACKNOWLEDGMENTS

Financial disclosure: Funding provided by a grant from the National Cancer Institute, CA118953 (P.I. S. Lee)

Declaration of Competing Interest: There are no conflicts of interest to report.

REFERENCES

1. Flowers ME, Parker PM, Johnston LJ, et al. Comparison of chronic graft-versus-host disease after transplantation of peripheral blood stem cells versus bone marrow in allogeneic recipients: long-term follow-up of a randomized trial. *Blood*. 2002;100:415–419.
2. Zecca M, Prete A, Rondelli R, et al. Chronic graft-versus-host disease in children: incidence, risk factors, and impact on outcome. *Blood*. 2002;100:1192–1200.
3. 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.
4. Flowers ME, Storer B, Carpenter P, et al. Treatment change as a predictor of outcome among patients with classic chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2008;14:1380–1384.
5. Martin PJ, Storer BE, Inamoto Y, et al. An endpoint associated with clinical benefit after initial treatment of chronic graft-versus-host disease. *Blood*. 2017;130:360–367.

6. 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;11:945–956.
7. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. The 2014 Diagnosis and Staging Working Group report. *Biol Blood Marrow Transplant*. 2015;21:389–401 e1.
8. Lee SJ, Wolff D, Kitko C, et al. Measuring therapeutic response in chronic graft-versus-host disease. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: IV. The 2014 Response Criteria Working Group report. *Biol Blood Marrow Transplant*. 2015;21:984–999.
9. Pavletic SZ, Martin P, Lee SJ, et al. Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: IV. Response Criteria Working Group report. *Biol Blood Marrow Transplant*. 2006;12:252–266.
10. Lee S, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2002;8:444–452.
11. Palmer J, Chai X, Martin PJ, et al. Failure-free survival in a prospective cohort of patients with chronic graft-versus-host disease. *Haematologica*. 2015;100:690–695.
12. Palmer J, Chai X, Pidala J, et al. Predictors of survival, nonrelapse mortality, and failure-free survival in patients treated for chronic graft-versus-host disease. *Blood*. 2016;127:160–166.