Biol Blood Marrow Transplant 25 (2019) 1682–1688

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

Assessment of Individual versus Composite Endpoints of Acute Graft-versus-Host Disease in Determining Long-Term Survival after Allogeneic Transplantation

John Magenau[1,*], Thomas Braun[2], Erin Gatza[3], Tracey Churay[3], Amanda Mazzoli[3], Grant Chappell[3], Joseph Brisson[3], Lyndsey Runaas[4], Sarah Anand[1], Monalisa Ghosh[1], Mary Riwes[1], Attaphol Pawarode[1], Gregory Yanik[3], Pavan Reddy[1], Sung Won Choi[3]

[1] *Department of Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, Michigan*
[2] *Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan*
[3] *Department of Pediatrics, Michigan Medicine, University of Michigan, Ann Arbor, Michigan*
[4] *Department of Internal Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin*

Article history:
Received 21 November 2018
Accepted 22 January 2019

Keywords:
GVHD-free relapse-free survival
GRFS
GVHD
Composite endpoints

A B S T R A C T

The overall composite of graft-versus-host disease (GVHD)-free, relapse-free survival (GRFS), defined as survival free of grade III-IV acute GVHD (aGVHD), chronic GVHD (cGVHD) requiring systemic immunosuppressive therapy (IST), or relapse, has emerged as a useful composite in clinical trials and to capture clinically meaningful events that impact quantity and quality of survival after allogeneic hematopoietic cell transplantation (HCT). We reviewed 565 consecutive patients aged ≥18 years undergoing HCT for hematologic malignancy to analyze how baseline incidence, specifics of clinical definitions, and proposed reductions in any one individual event may dynamically alter the overall performance of the composite To determine the relative impact of each GRFS event (excluding death), we accounted for competing risks using Fine and Gray methods, and correlated each event with overall survival (OS) using Kaplan-Meier methods. The consequences of modulating individual or composite endpoints on OS, such as hypothesized reductions of events of an HCT interventional trial, were examined using Monte Carlo simulations. The median age of the cohort was 54 years (range, 18 to 73 years). The majority of patients received HLA-matched unrelated donor HCT (53%), consisting of peripheral blood stem cell grafts (90%) after myeloablative conditioning (68%). Relapse conferred the greatest risk for death (hazard ratio [HR], 7.89; 95% confidence interval [CI], 5.83 to 10.69), followed by grade III-IV aGVHD (HR, 6.16; 95% CI, 4.42 to 8.56) and cGVHD requiring IST (HR, 1.69; 95% CI, 1.16 to 2.46). The overall GRFS composite correlated with an HR of 4.81 (95% CI, 3.61 to 6.41), which was lower compared with either relapse or grade III-IV aGVHD. Statistical simulations found that modulating the combined risk of both relapse and grade III-IV aGVHD predicted the greatest change in 5-year OS. These simulations suggest that GRFS as currently defined may be less optimal for correlating with OS, and further refinement of composite endpoints is needed. Nonetheless, composite endpoints may be particularly helpful in mitigating potential difficulties in interpretation when competing risks are present, most commonly seen in HCT studies.

© 2019 Published by Elsevier Inc. on behalf of American Society for Blood and Marrow Transplantation.

INTRODUCTION

Allogeneic hematopoietic cell transplantation (HCT) is used worldwide as an accepted treatment approach to improve survival in many advanced hematologic malignancies [1]. Unfortunately, graft-versus-host disease (GVHD) remains a principle cause of nonrelapse mortality (NRM) and overall morbidity after allogeneic HCT [2]. Historically, the majority of acute GVHD (aGVHD) prevention studies have focused on limiting grade II-IV aGVHD as the primary endpoint along with relapse and NRM as competing risks [3]. However, the tight linkage between GVHD and graft-versus-leukemia (GVL) responses that also influence relapse, and by extension eventual overall survival (OS), remain important considerations when interpreting the overall clinical trial results [4]. Thus, although determining the impact of a particular prophylaxis strategy on aGVHD as a primary endpoint is essential, the success or failure of the strategy [5] may not be accurately captured with a single endpoint given the possibility of an intervention to

Financial disclosure: See Acknowledgments on page 1687.
The findings of this work were presented as an oral abstract at the Transplantation and Cellular Therapy Meeting of the American Society for Blood and Marrow Transplantation and Center for International Blood and Marrow Transplant Research, Houston, Texas, February 24, 2019.
* Correspondence and reprint requests: John Magenau, MD, 1500 E. Medical Center Drive, UH-South BMT Program, Ann Arbor, MI 48109.
E-mail address: johmage@med.umich.edu (J. Magenau).

https://doi.org/10.1016/j.bbmt.2019.01.024
1083-8791/© 2019 Published by Elsevier Inc. on behalf of American Society for Blood and Marrow Transplantation.

unintentionally ameliorate beneficial immunity. In recent years, there has been a focus on novel endpoints for GVHD studies, including composite measures that attempt to capture multiple clinically meaningful (and potentially interrelated) outcomes [6-8]. Composite endpoints are assumed to better estimate the likelihood of experiencing an ideal post-HCT clinical course by identifying factors that influence outcomes and in some contexts are supplanting historical endpoints in assessing primary outcome [8].

The Blood and Marrow Transplant (BMT) Clinical Trials Network recently proposed the composite endpoint of GVHD-free, relapse-free survival (GRFS), defined as the time from HCT to the first event of grade III-IV aGVHD, chronic GVHD (cGVHD) requiring systemic immunosuppressive therapy (IST), relapse, or death from any cause, as a more comprehensive approach of capturing the effectiveness of a given GVHD prevention strategy [6-8]. This analysis was based on previous single-center and registry analyses in large adult and pediatric HCT cohorts receiving GVHD prophylaxis with tacrolimus and methotrexate that estimated a 1-year GRFS in the range of 23% to 31%, suggesting that a composite endpoint identifies a large fraction of HCT survivors with suboptimal outcomes [6]. Subsequent studies have identified various patient and donor graft characteristics (eg, unrelated donor, peripheral blood stem cell [PBSC] graft) that are major determinants of GRFS [6,9-12], More recent studies have focused on refining the definitions of GRFS, particularly as related to the cGVHD component [13-16]. However, to our knowledge, no previous study has assessed the long-term survival implications of attaining the GRFS state when compared to traditional GVHD endpoints; and importantly, the degree to which altering GRFS, as hypothesized in HCT clinical trials, can modify a patient's subsequent risk of death.

In the present study, GRFS was analyzed for its importance as a prognostic measure by examining the relative ability of each individual clinical component and the overall composite in determining OS at 5-years post-HCT. Importantly, we determined that clinical components of GRFS (excluding survival) have key differences in their timing and relationship to OS. Using statistical simulations, we examined the relative consequences of successfully modifying individual or composite GVHD endpoints on long-term OS to further elucidate the role of GRFS in clinical trials research.

METHODS

Study Design

A retrospective cohort study was conducted on 565 consecutive patients aged ≥18 years undergoing first-time allogeneic HCT for hematologic malignancy between January 2007 and March 2013. The study was approved by the University of Michigan Institutional Review Board (IRB# HUM00043287). Data abstraction of patient-, disease-, and transplantation-related variables was performed through manual chart review of the electronic medical record system (CareWeb and MiChart/EPIC), supported by the University of Michigan Electronic Medical Record Search Engine (EMERSE). EMERSE is designed to comprehensively search all institutional clinical documents using specified search terms and queries [17]. Documents screened by EMERSE were examined in more detail for relevant study data. Outcomes of the study included aGVHD, cGVHD, relapse, NRM, and OS. GRFS events included grade III-IV aGVHD, cGVHD requiring IST, relapse, or death from any cause during the first year after allogeneic HCT. Importantly, as originally defined, death was a component of GRFS [6-8]. In analyses where clinical events comprising the GRFS composite were assessed for their long-term impact on OS, early deaths (ie, occurring prior to the endpoint[s] of interest) are necessarily excluded. In these instances, the composite of grade III-IV aGVHD, cGVHD requiring IST, and relapse (excluding death) more closely reflected "GRF" (excluding survival).

Patient-, Disease-, and Transplantation-Related Characteristics

Demographic data, including age at transplantation, sex, race, ethnicity, disease risk (low, intermediate, high) [18], and HCT comorbidity index (HCT-CI) scores (low, intermediate, high) [19], were collected for each patient and recorded in the integrated University of Michigan BMT Clinical Research Database and Biorepository. Transplantation-related characteristics included conditioning regimen, conditioning intensity [20], stem cell source (bone marrow, peripheral blood, or cord blood), donor and recipient characteristics (sex, related or unrelated, HLA-matched [6/6, 8/8, or 10/10] or -mismatched [4/6, 5/6, 7/8, or 9/10], ABO and Rh blood types), and number of $CD34^+$ cells transfused. GVHD prophylaxis regimen was also recorded.

Statistical Analysis

Statistical analyses were performed using R version 3.4.0 (R Institute for Statistical Computing, Vienna, Austria), with $\alpha = .05$ defining the level of statistical significance (2-sided). Summary data were calculated for patient-, disease-, and transplantation-related variables, with medians and ranges determined for continuous variables and counts and percentages calculated for categorical variables. We estimated the cumulative incidence for relapse, aGVHD, and cGVHD, accounting for competing risks using the methods of Fine and Gray [3]. OS was estimated with Kaplan-Meier methods [21]. The endpoint for GRFS was the earliest of grade III-IV aGVHD, cGVHD requiring IST, relapse, death, or last follow-up. GRFS rates were estimated using Kaplan-Meier methods, and Cox regression was used to model the association of GRFS with patient-, disease-, and transplantation-related variables. Multivariate regression models using stepwise regression selection was used to identify the best-fitting models for the GRFS outcome. The following variables were included: age at transplantation, donor, graft source, year of HCT (ie, treating each year 2007 to 2013 as a group), disease risk, HCT-CI, and conditioning intensity. The associations of aGVHD, relapse, cGVHD, and their composite with OS was modeled by treating each outcome as a time-varying covariate in a Cox regression model.

Simulations were performed to assess the degree of improvement in OS would be expected from reductions of individual GRFS events and of the GRFS composite. Using the joint distribution of aGVHD, relapse, cGVHD, NRM, and follow-up seen in our data, we simulated the outcomes of 500 hypothetical patients as a new realization of our data, with the exception that we reduced the rates of 1 of the outcomes. From these data, we estimated the observed 5-year survival rate. We repeated this process for a total of 5000 simulations, and here we report the median 5-year OS rate among those 5000 simulations.

RESULTS

Patient and Transplantation Characteristics

The median age of the study population was 54 years (range, 18 to 73 years; interquartile range, 45 to 60 years), composed predominantly of males (58%) and Caucasians (94%). The most common indication for transplantation was acute myelogenous leukemia (41%). HLA-matched unrelated donor (MUD) HCT (n = 301; 53%) was the most frequent type, and peripheral blood (90%) was the predominant stem cell source. Three hundred eighty-four patients (68%) were conditioned with a myeloablative regimen (including total body irradiation ≥1200 cGy in 16% of these), according to the American Society for Blood and Marrow Transplantation working definitions [20]. Only 6 patients (1%) were conditioned with regimens that involved anti-human T cell antibody (antithymocyte globulin/thymoglobulin or alemtuzumab), and no patient received ex vivo T cell depletion. $CD34^+$ cells were infused at a median concentration of 5.8×10^6 cells/kg (range, $.12 \times 10^6$ to 1.4×10^7; interquartile range. 4.6×10^6 to 6.9×10^6). GVHD prophylaxis included either tacrolimus-methotrexate (n = 286; 51%) or tacrolimus-mycophenolate mofetil (n = 260; 46%) in the majority of patients (Table 1).

Incidence of GRFS and GVHD Endpoints

The cumulative incidences of grade II-IV aGVHD and grade III-IV aGVHD by day 180 were 40% (95% CI, 36% to 44%) and 17% (95% CI, 14% to 20%), respectively. As expected, the majority of aGVHD events occurred early, ≥90% by day +180. The median time to onset of cGVHD was 173 days, and the 1-year incidence of cGVHD requiring IST was 45% (95% CI, 41% to 50%). The incidence of relapse at 1 year was 32% (95% CI, 28% to 36%), with the majority of relapses occurring within 6 months. Overall, only 21% of patients achieved GRFS, defined as being alive and not having experienced grade III-IV aGVHD, cGVHD requiring IST, or relapse at 1-year post-HCT. In multiple regression analysis, risk factors independently associated with GRFS included

Table 1
Patient Demographic Data and Transplant Characteristics

Parameter Demographics	Value
Number of patients	565
Recipient age, yr, median (range)	54 (18-73)
Sex, n (%)	
Male	330 (58)
Female	235 (42)
Race/ethnicity, n (%)	
Caucasian	532 (94)
African American/black	12 (2)
Asian	8 (1)
Other	13 (2)
Transplantation characteristics	
Diagnosis, n (%)	
Acute myelogenous leukemia	233 (41)
Acute lymphocytic leukemia	67 (12)
Lymphoma (Hodgkin, non-Hodgkin)	117 (21)
Myelodysplastic syndrome/myeloproliferative disease	81 (14)
Plasma cell (leukemia, myeloma, other)	44 (8)
Other	23 (4)
Disease status, n (%)	
Low	217 (38)
Intermediate	138 (24)
High	188 (33)
Other	22 (4)
HCT-CI, n (%)	
Low	194 (34)
Intermediate	186 (33)
High	185 (33)
Conditioning intensity, n (%)	
Full	384 (68)
Reduced	181 (32)
Donor type, n (%)	
Matched related	250 (44)
Matched unrelated	217 (38)
Mismatched related	14 (2)
Mismatched unrelated	84 (15)
Cell source, n (%)	
PBSCs	507 (90)
Bone marrow	32 (6)
Cord blood	25 (4)
$CD34^+$ cell count, 10^6 cells/kg, median (range)	5.8 (.12-13.6)
Sex mismatch (donor-recipient), n (%)	
Female-male	111 (20)
Male-female	145 (26)
GVHD prophylaxis regimen, n (%)	
Tacrolimus/methotrexate	286 (51)
Tacrolimus/mycophenolate mofetil*	260 (46)
Tacrolimus/sirolimus†	18 (3)
Tacrolimus	1 (.2)

* Four patients received cyclosporine A/mycophenolate mofetil.
† One patient received tacrolimus/sirolimus/methotrexate.

disease risk (HR, 1.63; 95% CI, 1.31 to 2.04) and year of HCT (HR, 1.36; 95% CI, .99 to 1.88). Across the entire cohort, relapse and cGVHD requiring IST were the most common events leading to GRFS at 1 year (Supplementary Figure 1). Donor type was not associated with GRFS in univariate or multiple regression analyses. Relapse was the most common event leading to GRFS after HCT from an HLA-matched related donor (47% versus 27%), whereas cGVHD requiring IST was the most common event after MUD HCT. Relapse was the leading cause of death in the cohort, followed by cGVHD and aGVHD (Supplementary Table 1). The 180-day and 1-year incidences of NRM and OS were 12% and 16% and 74% and 61%, respectively (Table 2).

Assessing Impact of Individual GRFS Components on OS

To determine the contribution of each GRFS event, we correlated each event with OS. Although not a GRFS event, grade II-IV aGVHD is a standard endpoint for GVHD prevention trials.

Table 2
Clinical Outcomes

Outcome	Days Post-HCT	Estimate, % (95% CI)
Grade II-IV aGVHD	100	34 (31-38)
	180	40 (36-44)
	365	41 (37-45)
Grade III-IV aGVHD	100	15 (12-18)
	180	17 (14-20)
	365	17 (14-20)
Relapse	180	27 (24-32)
	365	32 (28-36)
cGVHD	180	25 (21-28)
	365	45 (41-50)
NRM	180	12 (9-15)
	365	16 (14-20)
OS	180	74 (70-77)
	365	60 (57-65)
GRFS composite	180	30 (26-34)
	365	21 (18-25)

We found that it was associated with an HR of 1.71 for death (95% CI, 1.39 to 2.10). The risk was not seen with grade II aGVHD alone (HR, 1.09; 95% CI, .86 to 1.39). As expected, severe grade III-IV aGVHD best correlated with increased risk for death (HR, 6.16; 95% CI, 4.42 to 8.56). Relapse conferred the greatest risk for death compared with any other event (HR 7.89; 95% CI, 5.83 to 9.69) (Table 3).

cGVHD requiring any IST has been incorporated into composite endpoints, in recognition of the risk of morbidity and mortality following HCT [6,9,10,12,13,15,16,22-27]. In our cohort, we identified cGVHD requiring IST as an independent risk factor for death (HR, 1.69; 95% CI, 1.16 to 2.46) (Table 3). Because systemic corticosteroid use could reflect more severe cGVHD, we specifically examined this subset of patients with cGVHD for its relationship with OS. Indeed, patients with cGVHD requiring systemic corticosteroids had a greater risk of death (HR, 2.07; 95% CI, 1.42 to 3.00) compared with those receiving IST without corticosteroids (HR, 1.22; 95% CI, .69 to 2.16) (Table 4).

Taking grade III-IV aGVHD, cGVHD requiring IST, or relapse into account as a GRFS composite endpoint (HR, 4.81; 95% CI, 3.61 to 6.41), the risk for death was notably greater than that associated with cGVHD requiring IST alone (HR, 1.69; 95% CI, 1.16 to 2.46) or the traditional endpoint of grade II-IV aGVHD alone (HR, 1.09; 95% CI, .86 to 1.39). Interestingly, the GRFS composite was correlated with a lower risk for death compared with either grade III-IV aGVHD (HR, 6.16; 95% CI, 4.42 to 8.56) or relapse alone (HR, 7.89; 95% CI, 5.83 to 10.69) (Table 3). These data suggest that individual components of the GRFS composite, and the manner in which they are defined, contribute differentially to the risk of death.

Comparison of the GRFS Composite on OS

Although various pre-HCT factors can predict GRFS, the importance of the GRFS composite and its association with long-term OS remain uncertain. As expected, GRFS, as

Table 3
Association of Clinical Outcomes with Death after HCT

Outcome	HR (95% CI)
Grade II aGVHD	1.09 (.86-1.39)
Grade II-IV aGVHD	1.71 (1.39-2.10)
Grade III-IV aGVHD	6.16 (4.42-8.56)
Relapse	7.89 (5.83-10.69)
cGVHD requiring IST	1.69 (1.16-2.46)
GRFS composite	4.81 (3.61-6.41)

Table 4
Association of cGVHD with Death after HCT

Outcome	cGVHD Treatment	HR (95% CI)
cGVHD	Any IST	1.69 (1.16-2.46)
cGVHD	Steroids only	2.07 (1.42-3.00)
cGVHD	IST, no steroids	1.22 (.69-2.16)

originally defined to include those with cGVHD requiring any IST [6,28,29], was 21% (95% CI, 18% to 25%), and was associated with an HR of 4.81 for death (95% CI, 3.61 to 6.41). Recognizing that the overall risk for death could be dependent on how cGVHD treatment (eg, IST with or without corticosteroids) was defined within the composite, various GRFS measures were analyzed for their relative impacts on mortality (Table 5). GRFS that included patients with cGVHD requiring corticosteroids (HR, 5.12; 95% CI, 3.75 to 7.00) did not differ from that including patients with cGVHD requiring IST but no corticosteroids (HR, 5.01; 95% CI, 3.94 to 6.39) (Table 5). Given that cGVHD had the lowest association with death (versus relapse or grade III-IV aGVHD), we modified GRFS to remove cGVHD for its impact on OS. This modified GRFS was associated with a greater risk for death (HR, 6.65; 95% CI, 5.24 to 8.43) compared with the conventional GRFS composite (HR, 4.81; 95% CI, 3.61 to 6.41) (Table 6).

Effect of Modulating Endpoints on Survival: Monte Carlo Simulations

To determine the dynamic impact of successfully altering an individual clinical event or a composite endpoint (eg, GRFS), Monte Carlo simulations were performed to model the impact of hypothetical improvements on individual events or combined events (composite) on OS. In a hypothetical simulation, reducing grade II to IV aGVHD from a historical incidence of 44% at day +180 to a hypothesized incidence of 33% or 22% (ie, relative reduction of 25% or 50%, respectively), assuming no change in relapse incidence, had a negligible impact on OS. For example, reducing the incidence of grade II-IV aGVHD by 50% predicted only a 2% absolute increase in 5-year OS, from 59% to 61%. When we simulated a 50% relative reduction in grade III-IV aGVHD, there again was no major impact on OS at 5 years (Table 7). However, a 50% reduction in the incidence of relapse predicted a 5% absolute increase in 5-year OS.

To determine whether simultaneous alterations in composite endpoints that include greater numbers of potentially interrelated events could impact late OS, we simulated reductions in the day +180 incidence of modified GRFS and conventional GRFS events. In a stepwise approach, we simulated the effect of reducing the combined risk for the modified GRFS and conventional GRFS composite endpoints at day +180 on 5-year OS. Deaths from any cause before this timepoint were excluded, and we focused on the nonfatal clinical events included in GRFS. In the case of modified GRFS, patients' combined risk of grade III-IV aGVHD and relapse was reduced proportionally by 25% or 50%. For conventional GRFS, patients' combined risk of experiencing grade III-IV aGVHD, relapse, or cGVHD requiring any IST was reduced proportionally by 25% or 50%. These simulations suggest that hypothetical alterations in composite endpoints may produce meaningful changes in 5-year OS. For example, reducing the combined risk of grade III-IV aGVHD and relapse (modified GRFS) at day +180 by 50% predicted a 7% improvement in 5-year OS. Similarly, reducing the combined risk of all 3 component endpoints of conventional GRFS by 50% predicted a 6% improvement in 5-year OS (Table 7).

Table 5
Impact of cGVHD Definition on GRFS Endpoints and Association with Death at 1 Year after HCT

Outcome	cGVHD Treatment	Estimate, % (95% CI)	HR (95% CI)
GRFS	Any IST	21 (25-18)	4.81 (3.61-6.41)
GRFS	Steroids only	18 (21-15)	5.12 (3.75-7.00)
GRFS	IST, no steroids	33 (38-30)	5.01 (3.94-6.39)

Table 6
Time-Varying Analysis for Association of Clinical Events with Death after HCT

Outcome	Time-Varying HR (95% CI)
Grade II-IV aGVHD	1.71 (1.39-2.10)
Grade III-IV aGVHD	6.16 (4.42-8.56)
cGVHD (IST, steroids)	1.69 (1.16-2.46)
Relapse	7.89 (5.83-10.69)
GRFS composite	4.81 (3.61-6.41)
Modified GRFS composite*	6.65 (5.24-8.43)

* Grade III-IV aGVHD-free relapse-free survival (no cGVHD).

DISCUSSION

The GRFS composite as originally defined [6-8] has emerged as a comprehensive approach in aGVHD prevention studies and for capturing clinically meaningful events that characterize ideal outcomes in the first year post-HCT [6,7,28,29]. Our present analyses extend these observations to describe the critical importance of the GRFS composite in determining long-term OS that is superior to traditional grade II-IV aGVHD endpoints. By correlating individual GRFS events (excluding death) with OS, we are able to better understand the relative contribution of each event. Our findings suggest that cGVHD may be a less impactful predictor for OS compared with other GRFS events from the perspective of long-term survival. Through statistical simulations, we highlight the critical role of modifying the key early events of severe aGVHD and malignant relapse to meaningfully alter survival post-HCT. Indeed, relapse showed the strongest association with survival.

A key consideration in merging variables in time-to-event analyses is determining their relative time course, relatedness, and individual contribution to the overall clinical outcome of interest. Moreover, the overall prevalence of severe aGVHD events has declined in recent years [30]; thus, capturing other potentially related events also may be useful for designing adequately powered clinical trials. cGVHD, a common complication post-HCT and a principle contributor to GRFS (excluding death), has known associations with survival [31,32] but, perhaps more importantly, is recognized as a surrogate for health-related quality-of-life (HRQOL) [33]. Although cGVHD may negatively impact HRQOL, its clinical onset is delayed, tends to wax and wane over time, and may be protective against relapse [31]. These attributes differ inherently from the other temporally related early GRFS events (excluding death), which are significantly associated with mortality, suggesting that in some contexts, cGVHD may over-represent outcomes of a composite endpoint [16,26]. To address such concerns, several groups have attempted to redefine the cGVHD endpoint in a manner that identifies higher-risk disease states.

Current GRFS (cGRFS) was an attempt by Solomon et al [16] to dynamically estimate the probability of being alive, in remission, and without clinically significant cGVHD, as defined by the NIH consensus criteria, at any time post-HCT. In their study, moderate-to-severe cGVHD was observed to decrease over time. Compared with the conventional GRFS definition

Table 7
Monte Carlo Simulations

Event	Baseline % at Day 180	Relative % Change	Estimated OS, %	% Change in OS
Grade II-IV aGVHD	44	↓25	60	1
		↓50	61	2
Grade III-IV aGVHD	19	↓25	60	1
		↓50	61	2
Relapse	30	↓25	62	3
		↓50	64	5
Grade III-IV aGVHD or relapse	49	↓25	62	3
		↓50	66	7
Grade III-IV aGVHD, relapse, or cGVHD requiring IST	66	↓25	62	3
		↓50	65	6

[6,7,28,29], which was met by approximately 25% of patients, cGRFS will reclassify almost one-half of patients originally censored for an event, suggesting that revisions in the operation definition and timing of cGVHD will greatly impact the prevalence of this variable (event). Kawamura et al [26] further refined cGRFS by excluding GVHD that resolved and did not require systemic IST. This variation of the composite endpoint that incorporates treatment responsiveness, termed refractory GRFS (rGRFS), appeared to be comparable to cGRFS.

These analyses indicate that definitions that include systemic IST (of any type) for cGVHD capture a greater proportion of patients with less severe disease, thereby diminishing the overall effect on survival post-HCT. It is important to emphasize that the manner in which cGVHD is defined is likely to differ in various research or clinical settings, especially beyond 1-year post-HCT, when HRQOL may require increased emphasis because a greater proportion of patients will experience long-term survival. Systematic data capture of HRQOL measures, particularly in longitudinal survivorship studies, may provide new opportunities to integrate HRQOL into composite endpoints that more accurately capture the disease burden of cGVHD.

In this study, we have characterized a large, single-institution cohort with the conventional GRFS composite endpoint [6,8,29]. Our findings indicate that cGVHD requiring IST was associated with the lowest risk for death compared with relapse or grade III-IV aGVHD. In individuals who were censored for the cGVHD requiring IST event, the use of systemic corticosteroids influenced survival when treated independently but had a negligible impact when introduced as part of the overall composite, because there was no difference in mortality risk between individuals treated with corticosteroids only and those treated with IST but without corticosteroids. Thus, although modifying the definition of cGVHD treatment can greatly change the numbers of patients meeting the GRFS estimate (ie, corticosteroid treatment resulted in 18% GRFS, but systemic IST without but no corticosteroids resulted in 33% GRFS), it does not appear to greatly impact OS of the composite. The GRFS composite endpoint helps capture the complexity of outcomes that are important in allogeneic HCT (eg, morbidity, survival) by integrating important events relevant to the intervention.

A limitation of our statistical analyses is that we treated the hypothetical simulations equally across each event. However, we recognize that in real-world settings, the impact of an intervention on a composite endpoint does not necessarily imply that it is equal across individual components of the composite. Nonetheless, we report the individual outcomes herein as well as the composite GRFS endpoint. Future work should include further validation of our simulations in large, multicenter registry data (eg, Center for International Blood and Marrow Transplant Research), or in prospectively designed, multicenter clinical trials (BMT CTN).

In clinical trials where more immediate outcomes are of interest (eg, aGVHD prevention trials) or where a higher number of events occur early after allogeneic HCT, individual or composite endpoints that address the disease under study and/or tightly inter-related variables may be more clinically and biologically relevant. Because composite endpoints comprehensively capture greater numbers of interrelated events, they may also limit potential bias from competing risks. While composite endpoints help to estimate the effect of an intervention over a relatively short time span, they are also highly correlated with an ultimate event (eg, survival). GRFS in GVHD studies helps to capture events influenced by the intervention (eg, severe aGVHD, cGVHD, relapse, death). Despite heterogeneity of an intervention effect across these individual events of a composite endpoint (eg, GRFS) on an ultimate event (eg, survival), GRFS more accurately reflects the GVHD disease burden. Nonetheless, correlating early events (aGVHD) with survival versus estimating long-term morbidity (cGVHD) with survival may be different goals for different contexts (ie, acute vs chronic settings). Because GVHD trials are increasingly being designed with the GRFS composite endpoint, the changing landscape will continue to shape scientifically valid and clinically relevant endpoints, providing more insight into optimally utilizing these endpoints to conduct robust clinical trials.

In this study, grade III-IV aGVHD and relapse events were associated with the greatest risk of death in the first year post-HCT. These findings are consistent with the benchmark analysis performed by Pasquini et al [8], in which grade III-IV aGVHD was associated with HCT failure. When conventional GRFS was compared with a modified GRFS composite that included only grade III-IV aGVHD and relapse events (excluding cGVHD), the 2-variable composite produced a greater HR for survival. To further study this observation, we performed simulations that assessed the degree of improvement in variable(s) needed to change long-term OS. This was accomplished by modeling hypothetical reductions in GVHD, relapse, and the conventional GRFS composite. The greatest gains in survival were seen from proportionally reducing both grade III-IV aGVHD and relapse events (modified GRFS); however, including proportional reductions in cGVHD events requiring IST with grade III-IV aGVHD and relapse had no further effect on improving OS. These findings in no way diminish the importance of controlling cGVHD; rather, they highlight the prognostic importance of targeting major temporally related early post-HCT adverse events (ie, grade III-IV aGVHD and relapse) in aGVHD prevention studies owing to their strong connection to mortality.

Since the initial report by Holtan et al [6], more data are emerging from retrospective single- and multi-center studies [6,9,10,12,13,15,16,22-27] investigating the impact of GRFS on the relative success or failure after allogeneic HCT [5]. The

findings raise important questions regarding the use of GRFS in clinical trials given how various definitions and historical (or published) incidences of events in the population affect the findings. Solomon et al [16] and Kawamura et al [26], both show that individual GRFS events, and the manner in which they are defined, contribute differentially to the overall incidence of the composite. Further, as reported herein, the individual event of cGVHD was overrepresented in the composite given the baseline characteristics of our study population (ie, significant use of unrelated donor and PBSCs). Nonetheless, the general overlap in the distributions of relapse and aGVHD, their potential biological relatedness, and their accounting for 63% of all deaths in this study population support the use of their combination as an endpoint. Certainly, our simulations highlight a potential strength of composites over the traditional grade II-IV aGVHD in terms of impact on OS. However, recognizing the complexities of composites and the tight GVHD-GVL link [4], established single endpoints likely will remain key benchmarks given their straightforward readouts for GVHD. Furthermore, our findings raise some interesting questions, such as whether single interventions targeting GVHD (or relapse) can realistically impact a composite endpoint. For example, this study suggests that multipronged or sequential stepwise approaches in targeting both GVHD and relapse may be necessary to ultimately influence survival.

We recognize a number of limitations of our study, primarily related to its single-institution and retrospective design. This underscores the potential pitfalls of comparing conventional GRFS across studies due to different patient-, disease-, and transplantation-related characteristics. Our multivariate analyses identified only advanced disease and year of HCT as risk factors for GRFS, as also reported by Holtan et al [6]. Nonetheless, other variables identified as risk factors for a GRFS event, such as stem cell source (eg, PBSCs) and type of donor (eg, MUD), were not observed in our cohort [6,12,15]. This likely reflects the fact that our study population comprised predominantly T cell-replete PBSC donors, which limited our ability to determine the effect of graft source on GRFS events. It is certainly plausible that datasets enriched for T cell-depleted grafts or post-transplantation cyclophosphamide would have lower rates of cGVHD that would influence the findings. Interestingly, despite the heterogeneity of patient-, disease-, and transplantation-related variables across studies, relapse and GVHD remain major obstacles to successful transplantation, and <25% of patients achieve conventional GRFS at 1 year [8], as also observed in the present study. The composite outcome, and the simulations that we performed, demonstrate where the greatest impact on survival could be targeted early in the first year post-HCT. As we hypothesized, our simulations showed that grade III-IV aGVHD and relapse were the principle factors in determining OS. It is likely that further refinement of these endpoints will depend on the a priori defined study populations (eg, hematologic malignancies, acute myelogenous leukemia), as well as on the interventions provided (eg, aGVHD prevention, cGVHD treatment).

ACKNOWLEDGMENTS

The authors thank the patients who participated in the University of Michigan Long-Term Evaluation of the Biology and Outcomes of Hematopoietic Stem Cell Transplantation and Cellular Therapy (HUM43287) protocol. We would also like to thank Dr. Brian Parkin for his contributions to this manuscript.

Financial disclosure: This work was supported by the A. Alfred Taubman Medical Research Institute and the Edith S. Briskin/Shirley K Schlafer Foundation (S.W.C.). Dr. John Magenau is supported by an NIH/NIAID Career Development Award (K23 A123505-01).

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

Authorship statement: J.M.: conceptualization, methodology, writing-original draft, supervision, validation, writing-review/editing; T.B.: methodology, formal analysis, visualization, writing-review/editing; E.J.: data curation, visualization, writing-review/editing; T.C.: data curation, visualization, writing-review/editing; A.M.: data curation, visualization, writing-review/editing; G.C.: data curation, writing-review/editing; J.B.: data curation, writing-review/editing; L.R.: data curation, writing-review/editing; S.A.: writing-review/editing; M.G.: writing-review/editing; B.P.: writing-review/editing; A.P.: writing-review/editing; M.R.: writing-review/editing; G.Y.: writing-review/editing; P.R.: writing-review/editing; S.W.C.: data curation, investigation, methodology, formal analysis, resources, supervision, validation, visualization, writing-original draft, writing-review/editing.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi: 10.1016/j.bbmt.2019.01.024.

REFERENCES

1. Gratwohl A, Baldomero H, Aljurf M, et al. Hematopoietic stem cell transplantation: a global perspective. *JAMA*. 2010;303:1617–1624.
2. D'Souza A, Fretham C. Current uses and outcomes of hematopoietic cell transplantation (HCT): CIBMTR summary slides; 2018. Available at: http://www.cibmtr.org. Accessed 3 July 2019 .
3. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.
4. Negrin RS. Graft-versus-host disease versus graft-versus-leukemia. *Hematology Am Soc Hematol Educ Program*. 2015;2015:225–230.
5. Holtan SG. The perfect transplant. *Biol Blood Marrow Transplant*. 2017;23:1044–1045.
6. Holtan SG, DeFor TE, Lazaryan A, et al. Composite end point of graft-versus-host disease-free, relapse-free survival after allogeneic hematopoietic cell transplantation. *Blood*. 2015;125:1333–1338.
7. Sankoh AJ, Li H, D'Agostino Sr. RB. Use of composite endpoints in clinical trials. *Stat Med*. 2014;33:4709–4714.
8. Pasquini MC, Logan B, Jones RJ, et al. Blood and Marrow Transplant Clinical Trials Network report on the development of novel endpoints and selection of promising approaches for graft-versus-host disease prevention trials. *Biol Blood Marrow Transplant*. 2018;24:1274–1280.
9. Mehta RS, Peffault de Latour R, DeFor TE, et al. Improved graft-versus-host disease-free, relapse-free survival associated with bone marrow as the stem cell source in adults. *Haematologica*. 2016;101:764–772.
10. Konuma T, Kato S, Oiwa-Monna M, Ishii H, Tojo A, Takahashi S. Comparison of graft-versus-host disease-free, relapse-free survival of transplantation using matched sibling donor, matched unrelated donor or unrelated cord blood after myeloablative conditioning for adult patients with hematological malignancies. *Leuk Lymphoma*. 2016;57:2126–2132.
11. Liu YC, Chien SH, Fan NW, et al. Prognostic factors on the graft-versus-host disease-free and relapse-free survival after adult allogeneic hematopoietic stem cell transplantation. *Stem Cells Int*. 2016;2016 5143071.
12. Solh M, Zhang X, Connor K, et al. Factors predicting graft-versus-host disease-free, relapse-free survival after allogeneic hematopoietic cell transplantation: multivariable analysis from a single center. *Biol Blood Marrow Transplant*. 2016;22:1403–1409.
13. Battipaglia G, Ruggeri A, Labopin M, et al. Refined graft-versus-host disease/relapse-free survival in transplant from HLA-identical related or unrelated donors in acute myeloid leukemia. *Bone Marrow Transplant*. 2018;53:1295–1303.
14. Ruggeri A, Labopin M, Ciceri F, Mohty M, Nagler A. Definition of GVHD-free, relapse-free survival for registry-based studies: an ALWP-EBMT analysis on patients with AML in remission. *Bone Marrow Transplant*. 2016;51:610–611.
15. Solh M, Zhang X, Connor K, et al. Donor type and disease risk predict the success of allogeneic hematopoietic cell transplantation: a single-center analysis of 613 adult hematopoietic cell transplantation recipients using a modified composite endpoint. *Biol Blood Marrow Transplant*. 2017;23:2192–2198.
16. Solomon SR, Sizemore C, Zhang X, et al. Current graft-versus-host disease-free, relapse-free survival: a dynamic endpoint to better define efficacy after allogenic transplant. *Biol Blood Marrow Transplant*. 2017;23:1208–1214.
17. Hanauer DA. EMERSE: The Electronic Medical Record Search Engine. *AMIA Annu Symp Proc*. 2006;941.

18. American Society for Blood and Marrow Transplantation (ASBMT). RFI disease classifications and corresponding CIBMTR classifications, 2019. Available at: https://www.asbmt.org/practice-resources/rfi-forms. Accessed 3 July 2019.
19. Sorror ML, Maris MB, Storb R, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood*. 2005;106:2912–2919.
20. Bacigalupo A, Ballen K, Rizzo D, et al. Defining the intensity of conditioning regimens: working definitions. *Biol Blood Marrow Transplant*. 2009;15:1628–1633.
21. Kaplan EL, Meier P. Nonparametric-estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457–481.
22. Inagaki J, Fukano R, Noguchi M, Okamura J. A single-center analysis of chronic graft-versus-host disease-free, relapse-free survival after alternative donor stem cell transplantation in children with hematological malignancies. *Int J Hematol*. 2017;105:676–685.
23. Simonetta F, Masouridi-Levrat S, Beauverd Y, et al. Partial T-cell depletion improves the composite endpoint graft-versus-host disease-free, relapse-free survival after allogeneic hematopoietic stem cell transplantation. *Leuk Lymphoma*. 2018;59:590–600.
24. Tan J, Wang Y, Yu SJ, Ma YY, Lei HY, Liu QF. Prognostic factors on graft-versus-host disease-free and relapse-free survival after allogeneic hematopoietic stem cell transplantation for adults with acute leukemia. *Leuk Res*. 2017;59:1–7.
25. Czerw T, Labopin M, Giebel S, et al. Anti-thymocyte globulin improves survival free from relapse and graft-versus-host disease after allogeneic peripheral blood stem cell transplantation in patients with Philadelphia-negative acute lymphoblastic leukemia: an analysis by the Acute Leukemia Working Party of the EBMT. *Cancer*. 2018;124:2523–2533.
26. Kawamura K, Nakasone H, Kurosawa S, et al. Refractory graft-versus-host disease-free, relapse-free survival as an accurate and easy-to-calculate endpoint to assess the long-term transplant success. *Biol Blood Marrow Transplant*. 2018;24:1521–1526.
27. Park SS, Jeon YW, Min GJ, et al. Graft-versus-host disease-free, relapse-free survival after allogeneic stem cell transplantation for myelodysplastic syndrome. *Biol Blood Marrow Transplant*. 2019;25:63–72.
28. Weisdorf D, Carter S, Confer D, Ferrara J, Horowitz M. Blood and Marrow Transplant Clinical Trials Network (BMT CTN): addressing unanswered questions. *Biol Blood Marrow Transplant*. 2007;13:257–262. [discussion: 255-256].
29. Bolanos-Meade J, Koreth J, Reshef R. A multi-center phase II trial randomizing novel approaches for graft-versus-host disease prevention compared to contemporary controls (BMT CTN 1203). Available at: https://www.clinicaltrials.gov/ct2/show/NCT02208037. Accessed 27 September 2018.
30. Gooley TA, Chien JW, Pergam SA, et al. Reduced mortality after allogeneic hematopoietic-cell transplantation. *N Engl J Med*. 2010;363:2091–2101.
31. Lee SJ, Kim HT, Ho VT, et al. Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant*. 2006;38:305–310.
32. Pidala J, Kurland B, Chai X, et al. Patient-reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. *Blood*. 2011;117:4651–4657.
33. Kurosawa S, Oshima K, Yamaguchi T, et al. Quality of life after allogeneic hematopoietic cell transplantation according to affected organ and severity of chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2017;23:1749–1758.