The best endpoint for acute GVHD treatment trials

Margaret L. MacMillan,[1,2] Todd E. DeFor,[1,3] and Daniel J. Weisdorf[1,4]

[1]Blood and Marrow Transplant Program, [2]Department of Pediatrics, [3]Biostatistics and Informatics Core, Masonic Cancer Center, and [4]Department of Medicine, University of Minnesota Medical School, Minneapolis

The optimal primary endpoint for acute graft-versus-host disease (GVHD) therapeutic trials has not been established. In a retrospective analysis, we examined the response of 864 patients who received prednisone 60 mg/m^2/d for 14 days, followed by an 8-week taper, as initial therapy for acute GVHD from 1990-2007 at the University of Minnesota. Patients received grafts of human leukocyte antigen–matched sibling bone marrow (BM) or peripheral blood (PB; n = 315), partially matched sibling BM or PB (n = 24), unrelated donor BM or PB (n = 313), single (n = 89) or double (n = 123) umbilical cord blood. Day 28 responses were similar to day 56 responses and better than day 14 responses in predicting transplantation-related mortality (TRM). In multiple regression analysis, patients with no response at day 28 were 2.78 times (95% CI, 2.17-3.56 times; *P* < .001) more likely to experience TRM before 2 years than patients with a response. Other factors associated with significantly worse 2-year TRM include older age, high-risk disease, severe GVHD, and partially matched related BM/PB. No other differences in response by donor source were observed. These data suggest that day 28 is the best early endpoint for acute GVHD therapeutic trials in predicting 2-year TRM. (*Blood*. 2010; 115(26):5412-5417)

Introduction

The optimal primary and early endpoint for prospective acute graft-versus-host disease (GVHD) therapeutic clinical trials has not been established. The endpoint not only must correlate with definitive hematopoietic cell transplantation (HCT) outcomes (ie, transplantation-related mortality [TRM] or survival) but also must identify patients unlikely to respond, allowing additional GVHD therapy to be initiated in a timely fashion. A variety of endpoints have been used in clinical trials, including overall or organ-specific response to therapy at specific, usually early, time points after its initiation, TRM, and survival.[1-6] Recently, a novel response endpoint, very good partial response (VGPR), was proposed as a practical endpoint including both diagnostic and functional criteria,[7] but it has yet to be tested.

In an attempt to determine the optimal endpoint for GVHD trials, we examined response (including VGPR) in 864 patients to standardized therapy with prednisone 60 mg/m^2/d for 14 days, followed by an 8-week taper, at days 14, 28, and 56 after initiation of treatment and correlated these responses to TRM.

Methods

Study design

Clinical and laboratory data were systematically and prospectively collected on all patients undergoing HCT and entered into the University of Minnesota Blood and Marrow Transplant Database. All HCT protocols were reviewed and approved by the Masonic Cancer Center Protocol Review Committee and Human Subjects Institutional Review Board at the University of Minnesota. All patients and/or guardians signed the informed consent approved by institutional review board in accordance with the Declaration of Helsinki.

Between January 1990 and December 2007, 3538 patients underwent HCT at the University of Minnesota, 2406 of whom received an allogeneic transplant. Of these patients, 1149 developed grade I to IV acute GVHD. Patients (n = 864) were treated with prednisone 60 mg/m^2 orally (or methylprednisolone 48 mg/m^2 intravenously) as initial therapy for acute GVHD and were the subjects for this analysis. At the time of analysis, patients had 0.8 to 17.1 years of follow-up (median, 6.9 years).

Patient and transplantation characteristics

Patient characteristics, including year of transplantation, recipient age, sex, cytomegalovirus (CMV) serostatus, and underlying diagnosis, are shown in Table 1. Median patient age was 32 years (range, 0.2-69 years) with 35% being younger than 18 years. Standard disease risk was defined as acute leukemia in first or second complete remission, chronic myelogenous leukemia in first chronic phase, myelodysplastic syndrome without excess blasts, or nonmalignant diseases. All other patients were considered high risk.

Transplantation characteristics, including donor type, preparative therapy, and GVHD prophylaxis are shown in Table 1. Related sibling donors and recipients were typed at antigen level for human leukocyte antigen (HLA)–A, -B, and -DRB1 unless adequate family typing was not available to determine haplotypes, in which case allele level DNA typing was performed. For unrelated donors (URDs), donors and recipients were typed for HLA-A and -B at antigen level and allele level typing at HLA-DRB1 until June 2005, when testing at the allele level for all loci was implemented. Prospective allele level typing for HLA-C was incorporated for all donor sources beginning in June 2004. HLA matching followed the definitions used by the Center for International Blood and Marrow Transplant Research.[8] For umbilical cord blood (UCB) transplants, patients and donors were typed for HLA-A and -B at the antigen level and for DRB1 at the allele level. HLA-DQ and -DP were not considered in URD or UCB donor selection.

Graft sources included HLA identical sibling bone marrow (BM) or peripheral blood stem cells (PBSCs; n = 315), HLA-mismatched sibling BM or PBSCs, (n = 24), well-matched[8] URD BM or PBSCs (n = 79), partially matched URD BM or PBSCs (n = 108), mismatched URD BM or PBSCs (n = 126), single UCB (n = 89) or double UCB (n = 123).

Table 1. Patient and transplantation characteristics

Factor	n (%)
Total no. of patients	864
Year of transplantation	
1990-1995	289 (33)
1996-2000	278 (32)
2001-2005	238 (28)
2006-2007	59 (7)
Age	
Younger than 10 y	193 (22)
10-17 y	108 (13)
18-35 y	191 (22)
36 y or older	372 (43)
Median (range)	32 (0.2-69)
Sex	
Male	530 (61)
Female	334 (39)
CMV serostatus	
Recipient negative, donor negative	340 (39)
Recipient negative, donor positive	107 (12)
Recipient positive	417 (48)
Disease	
ALL	141 (16)
AML	193 (22)
CML	179 (21)
MDS/MPD	72 (8)
NHL	63 (7)
Other malignancies	52 (7)
SAA/FA	55 (6)
Immune deficiency	25 (3)
Storage disorder	82 (9)
Disease risk*	
Standard	431 (50)
High	433 (50)
Donor type	
Matched sibling	315 (36)
Mismatched sibling	24 (3)
Well-matched URD	79 (9)
Partially matched URD	108 (13)
Mismatched URD	126 (15)
Single UCB	89 (10)
Double UCB	123 (14)
Conditioning	
Myeloablative	736 (85)
Nonmyeloablative	128 (15)
GVHD prophylaxis	
CSA/MTX	411 (48)
CSA/MMF	195 (23)
CSA/MP	98 (11)
T-cell depletion	112 (13)
MTX/ATG/prednisone	33 (4)
Tacrolimus/MTX	5 (< 1)
Tacrolimus	10 (1)

ALL indicates acute lymphoblastic leukemia; AML, acute myelogenous leukemia; CML, chronic myelogenous leukemia; MDS, myelodysplastic syndrome; MPD, myeloproliferative disease; NHL, non-Hodgkin lymphoma; SAA, severe aplastic anemia; FA, Fanconi anemia; CSA, cyclosporine A; MTX, methotrexate; MMF, mycophenolate mofetil; MP, methylprednisolone; and ATG, antithymocyte globulin.

*Standard risk indicates acute leukemia in CR1 or CR2, CML in first chronic phase, MDS without excess blasts, or nonmalignant diseases. High risk indicates all others.

Details of the preparative therapy and GVHD prophylaxis as well as supportive therapy techniques have been previously reported.[9-18] The majority (89.5%) of patients received a total body irradiation–based regimen, and 10.5% of patients received a chemotherapy alone regimen. GVHD prophylaxis differed among stem cells sources as shown in Table 1 and consisted of cyclosporine A (CSA)– or tacrolimus-based therapy in

83% of patients, T-cell depletion in 13% of patients, and methotrexate (MTX) alone in 4% of patients.

Diagnosis, staging, and grading of GVHD

Symptoms of acute GVHD were graded by standard clinical criteria,[19,20] modified to include upper gastrointestinal acute GVHD per the GVHD consensus conference.[21,22] Grade of GVHD refers to clinical (not histologic) grade throughout this report. Initial grade was calculated with the maximum stage in each organ within a 10-day window (-5 to $+5$ days) of initiation of steroid therapy. Real-time staging and grading of each organ was determined weekly by the attending physician, supported by laboratory and clinical information and histologic confirmation when possible. The grading scheme was consistent throughout the study period. Although all patients' GVHD diagnoses and maximum GVHD grades were retrospectively reviewed by the Acute GVHD Grading Committee (M.L.M. and D.J.W.), the overall grade used in this analysis was determined by a computer algorithm, incorporating all available clinical and pathologic GVHD organ staging data as originally and prospectively recorded, not modified by retrospective review. Responses at weekly or biweekly endpoints were determined by review of the prospectively recorded staging and grading data.

GVHD therapy

All patients received daily, thrice divided doses of prednisone 60 mg/m^2/d orally (or methylprednisolone intravenous equivalent, 48 mg/m^2) for 7 consecutive days, followed by daily, single dose, prednisone for 7 days as initial therapy for acute GVHD. Patients were maintained on therapeutic levels of CSA in 795 patients (93%) or tacrolimus in 15 patients (2%). In addition, patients with acute skin GVHD were treated with topical 0.1% triamcinolone cream or 1% hydrocortisone cream (for facial rash) 3 times daily. If a response to prednisone was observed, patients continued therapy with single daily dose prednisone 60 mg/m^2/d orally through day 14 and then commenced a taper of steroids over 8 weeks.[21]

Measurement of GVHD response to prednisone

Response to therapy was evaluated by the attending physician and prospectively recorded weekly in the University of Minnesota BM Transplant Database by determining the GVHD clinical stage score for each time point ($\pm$ 3 days).[23] Response was determined from the maximum acute GVHD stage and grade in each organ at day 14 ($\pm$ 7 days), 28 ($\pm$ 7 days), and 56 ($\pm$ 14 days) after prednisone treatment was initiated. Complete response (CR) was defined as the complete resolution of acute GVHD symptoms in all organs, without secondary GVHD therapy. Partial response (PR) was defined as improvement in GVHD stage in all initial GVHD target organs without complete resolution and without worsening in any other GVHD target organs, without secondary GVHD therapy. VGPR was defined as improvement in GVHD in all initial GVHD target organs, with maximum stage I involvement in one or more organs (except upper gastrointestinal tract), without secondary GVHD therapy. No response (NR) was defined as the same grade of GVHD or progression of GVHD in any organ or death, or the addition of secondary GVHD therapy. Progression was defined as worsening GVHD in at least 1 organ with or without amelioration in any organ. Steroid resistant acute GVHD was defined as progression of acute GVHD after 4 days of treatment with prednisone or no improvement of acute GVHD after 7 days of treatment with prednisone. Patients with steroid resistant GVHD were treated with secondary therapy and were considered to have NR. If patients experienced a flare of acute GVHD during the steroid taper and required therapy with a boost of steroids or the additional GVHD therapy, they were also considered to have NR.

Supportive care

Patients received antibiotic prophylaxis suitable for the functionally hyposplenic state accompanying GVHD. Broad-spectrum intravenous antibacterial and as indicated antifungal antimicrobials were used when patients developed fever. Patients received acyclovir prophylaxis if they

Table 2. GVHD organ stage at initiation of prednisone therapy

Organ	Stage, n (%)				
	0	1	2	3	4
Skin	165 (19)	111 (13)	224 (26)	357 (41)	7 (1)
Liver	797 (92)	22 (3)	27 (3)	13 (2)	5 (1)
Rectal	668 (77)	113 (13)	40 (5)	39 (5)	4 (< 1)
Upper GI tract	644 (75)	220 (25)			

GI indicates gastrointestinal.

were seropositive for herpes simplex virus and/or CMV. Oral trimethoprim-sulfamethoxazole was given for *Pneumocystis carinii* pneumonia prophylaxis. CMV seronegative recipients received CMV-safe (seronegative or filtered) blood products.

Statistical analysis

The κ statistic was used to measure agreement between response at days 14, 28, and 56.[24] McNemar test was used to test the proportions of response over time. Kaplan-Meier curves were used to estimate the probabilities of overall survival (OS) and cumulative incidence was used to estimate the probability of TRM and chronic GVHD through 2 years after treatment, treating relapse as a competing risk for TRM and non-GVHD death as a competing risk for chronic GVHD.[25,26] Cox regression analysis was used to asses the effect of response on OS and the Fine and Gray competing hazards regression analysis was used for TRM.[27,28] Likelihood ratios were calculated for each model containing the day 14, day 28, and day 56 response scores after adjusting for CMV serostatus (recipient negative/donor negative vs recipient negative/donor positive vs recipient positive), age (< 10 vs 10-17 vs 18-35 vs ≥36 years), disease risk (standard vs high), donor type (matched sibling vs URD well matched vs URD partially matched vs URD or sibling mismatched vs single UCB vs double UCB), conditioning (myeloablative [MA] vs nonmyeloablative [NMA]), initial acute GVHD grade at treatment initiation (I vs II versus III-IV), affected organs (skin only vs other), GVHD prophylaxis (MTX/antithymocyte globulin/prednisone vs T-cell depletion vs CSA-containing therapy) and days to treatment (< 28 vs ≥ 28). Although a likelihood ratio indicates that there is a trend toward better prediction of an outcome, there is no formal statistical test to compare these. A more formal statistical test was performed to compare the responses as predictors of TRM by computing an index of concordance, the C statistic (C). This statistic, derived by Harrel et al,[29] estimates the probability that, of 2 randomly chosen pairs of patients, the patient with the higher response will outlive the patient with the lower response. Values of C near 0.5 indicate that the response is no better than chance in determining which patient will live longer. Values of C near 1.0 indicate the response virtually always determines that the patient with the higher response has better survival. Certain pairs of observations were excluded from the calculation because of (1) pairs who had equivalent index response, (2) subsequent pairs who were both censored at the time of

Table 3. Response to GVHD therapy

Factors	n (%)
Day 14 response	
CR	306 (35)
VGPR	137 (16)
PR	65 (8)
NR	356 (41)
Day 28 response	
CR	461 (53)
VGPR	66 (8)
PR	36 (4)
NR*	301 (35)
Day 56 response	
CR	479 (55)
VGPR	43 (5)
PR	16 (2)
NR†	326 (38)

*Includes 32 patients who died before day 14.
†Includes 104 patients who died before day 42.

analysis, or (3) pairs in which one patient was censored before the event of the other patient. For TRM, patients were censored at the time of relapse or disease progression. The C statistic was computed for survival and TRM based on time to event analysis over the first 2 years after transplantation. Standard errors for the difference in the C statistics for the 2 index scores were estimated by applying a bootstrap procedure to the dataset with the use of 100 bootstrap samples.[30]

Results

Maximum initial GVHD stage in each organ is shown in Table 2. Initial GVHD organ involvement was skin only (n = 498; 57%), gut only (n = 146; 17%), liver only (n = 7; 1%), or multiorgan (n = 213; 25%). Before initiation of steroid therapy, initial GVHD grades were grade I in 230 patients (27%), grade II in 504 patients (58%), grade III in 119 patients (14%), and grade IV in 11 patients (1%). Median time to onset of GVHD from day of HCT was 32 days (range, 8-99 days). Median time to treatment with prednisone from day of HCT was 33 days (range, 8-99 days).

Response (CR, VGPR, PR, NR) at days 14, 28, and 56 after initiation of steroid therapy for acute GVHD is shown in Table 3 and Figure 1. Overall response (CR + VGPR + PR) was observed in a similar proportion of patients over time, being 59% at day 14, 65% at day 28, and 62% at day 56. However, a greater proportion of patients achieved CR at the later time points (53% at day 28, and
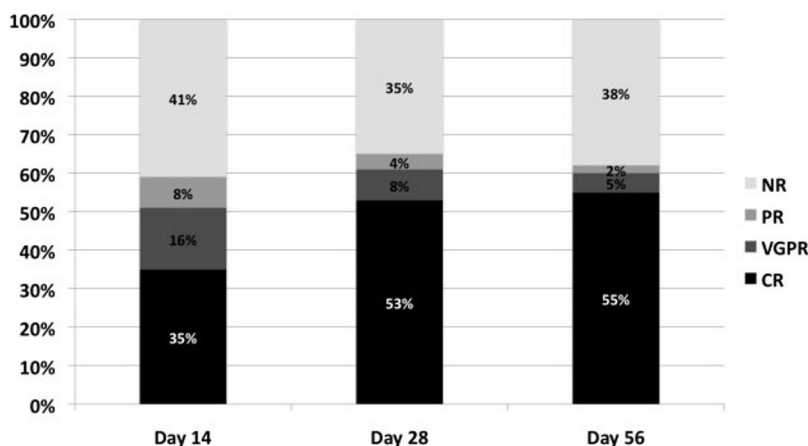
Figure 1. Response incidence (percentage of CR, VGPR, PR, NR) at day 14, day 28, and day 56 after initiation of steroid therapy for acute GVHD.

Table 4. Measure of agreement between treatment responses

	CR	VGPR	PR	NR	Total
Day 14 response compared with day 28 response*					
CR	276 (90)	5 (2)	2 (1)	23 (8)	306
VGPR	79 (58)	30 (22)	12 (9)	16 (12)	137
PR	26 (40)	17 (26)	11 (17)	11 (17)	65
NR	80 (22)	14 (4)	11 (3)	251 (70)	356
Total	461	66	36	301	864
Day 14 response compared with day 56 response†					
CR	246 (80)	4 (1)	4 (1)	52 (17)	306
VGPR	86 (63)	19 (14)	5 (4)	27 (19)	137
PR	38 (58)	9 (14)	4 (6)	14 (22)	65
NR	109 (31)	11 (3)	3 (1)	233 (65)	356
Total	479	43	16	326	864
Day 28 response compared with day 56 response‡					
CR	388 (84)	8 (2)	5 (1)	60 (13)	461
VGPR	36 (55)	20 (30)	1 (2)	9 (14)	66
PR	13 (36)	10 (28)	6 (17)	7 (19)	36
NR	42 (16)	5 (2)	4 (1)	250 (81)	301
Total	479	43	16	326	864

Values are no. of patients or n (%).
*Weighted κ = 0.58 (95% CI, 0.53-0.62), good agreement.
†Weighted κ = 0.42 (95% CI, 0.37-0.48), good agreement.
‡Weighted κ = 0.66 (95% CI, 0.61-0.70), good agreement.

55% at day 56) compared with day 14 (35%) ($P < .01$). In contrast, rates of VGPR (16%, 8%, 5%) ($P < .01$) and to less an extent PR (8%, 4%, 2%) ($P < .01$) decreased over time (day 14, 28, and 56, respectively), suggesting that patients who could respond had, for the most part, done so by day 28.

We then compared the incidence of response to steroid therapy by day of response, and agreement between treatment responses at each time point was measured (Table 4). Patients with PR at day 14 were more likely to have changed their responses by day 28 or 56. Day 28 responses were more stable and were similar to day 56 responses (κ = 0.66; 95% CI, 0.61-0.70), whereas there was less agreement to day 14 responses (κ = 0.58; 95% CI, 0.53-0.62). There was less agreement between day 14 and day 56 responses (κ = 0.42; 95% CI, 0.37-0.48). This confirms the improvement in response between day 14 and day 28, and the near concordance of response between day 28 and day 56. This result held true when VGPR response was included with PR (data not shown).

Chronic GVHD

For the entire cohort of 864 patients, the cumulative incidence of chronic GVHD at 2 years after initiation of steroid therapy was 41% (95% CI, 37%-45%). The incidence was highest in patients with PR at day 28 of steroids for acute GVHD (58%; 95% CI, 40%-77%) versus patients with CR (43%; 95% CI, 38%-48%), VGPR (41%; 95% CI, 28%-54%), or NR (41%; 95% CI, 34%-48%; $P \leq .001$).

Transplant-related mortality

For the entire cohort of 864 patients, TRM at 2 years after initiation of steroid therapy for GVHD was 36% (95% CI, 33%-40%). TRM at 2 years was highest for the 269 patients with NR at day 28 (52%; 95% CI, 46%-59%) compared with patients with a response (26%; 95% CI, 22%-30%) for 461 patients with CR, 18% (95% CI, 9%-28%) for 66 patients with VGPR, and 36% (95% CI, 19%-53%) for 36 patients with PR ($P < .001$) as shown in Figure 2.

We then analyzed the C statistic, an index of concordance, to measure the strength of association between response (CR, VGPR, PR, or NR) to GVHD therapy (at day 14, day 28, day 56) in

predicting 2-year TRM. Day 28 responses (C = 0.65) were slightly less predictive of TRM than day 56 responses (C = 0.75; $P < .001$) yet were better than day 14 responses (C = 0.57; $P < .001$). These findings held true when restricting the analysis to recipients of nonmyeloablative conditioning (n = 128) or those who received UCB transplants (n = 212; data not shown).

Because it is earlier, day 28 is a preferred time point over day 56 to promptly identify patients in need of additional GVHD therapy. Therefore, we used day 28 response as the endpoint in further analyses of factors associated with TRM. In multiple regression analysis, patients with NR at day 28 were 2.78 (95% CI, 2.17%-3.56%) times more likely to have 2-year TRM than patients achieving either CR, VGPR, or PR or patients with VGPR + PR ($P < .001$; Table 5) at day 28. Other factors associated with significantly worse 2-year TRM include severe grades III to IV GVHD at the time of initial steroid treatment (relative risk [RR], 1.63; 95% CI, 1.00-2.66; $P = .05$), skin-only GVHD (RR, 1.38; 95% CI, 1.01-1.89; $P = .04$), and a high-risk diagnosis at the time of HCT (RR, 1.36; 95% CI, 1.07-1.73; $P = .01$). In addition, older patients had significantly higher risk of TRM ($P < .01$). Two-year TRM was also significantly higher in recipients of partially matched URD (RR, 1.61; 95% CI, 1.09-2.38; $P = .02$) and mismatched URD or sibling donor (RR, 1.88; 95% CI, 1.34-2.63;
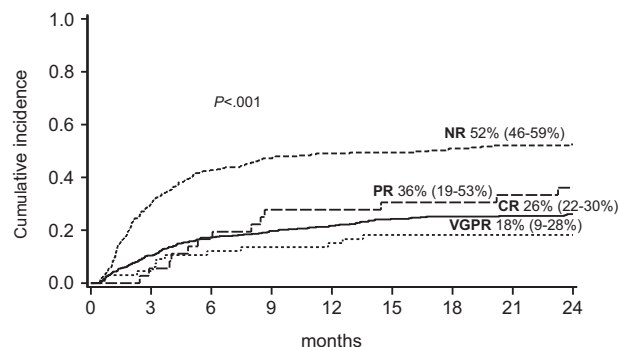
Figure 2. Cumulative incidence of TRM at 2 years by response at day 28 after initiation of steroid therapy for acute GVHD.

Table 5. Factors associated with 2-year TRM: multivariate analysis

Factor	Relative risk of TRM (95% CI)	P	Overall P
Day 28 response			< .001
CR*	1.0		
VGPR	0.61 (0.34-1.10)	.10	
PR	1.21 (0.71-2.06)	.48	
VGPR + PR	0.82 (0.54-1.25)	.36	
NR	2.78 (2.17-3.56)	< .001	
Initial acute GVHD grade			.05
I*	1.0		
II	1.13 (0.83-1.54)	.42	
III-IV	1.63 (1.00-2.66)	.05	
Organ involvement			.04
Other organs*	1.0		
Skin only	1.38 (1.01-1.89)		
Age			< .01
Younger than 10 y*	1.0		
10-17 y	1.65 (1.09-2.66)	.02	
18-35 y	1.69 (1.13-2.54)	.01	
36 y or older	1.94 (1.31-2.89)	.01	
Disease risk			.01
Standard*	1.0		
High	1.36 (1.07-1.73)	.01	
Donor type			< .01
Matched sibling*	1.0		
URD well-matched	1.42 (0.92-2.20)	.11	
URD partially matched	1.61 (1.09-2.38)	.02	
URD- or sibling-mismatched	1.88 (1.34-2.63)	< .001	
Single UCB	1.18 (0.72-1.94)	.52	
Double UCB	0.73 (0.44-1.20)	.21	

*Reference group.

$P < .001$). TRM was not affected by patient or donor CMV serostatus, conditioning intensity, GVHD prophylaxis, or days to steroid treatment.

To verify the effect of baseline patient and transplantation factors on 2-year TRM, we revised the multiple regression analysis and excluded day 28 response as a factor in the analysis. We again observed that factors associated with higher 2-year TRM included severe grade III to IV GVHD at the time of initial steroid treatment (RR, 2.04; 95% CI, 1.28-3.27; $P < .01$), skin-only GVHD (RR, 1.35; 95% CI, 1.00-1.83; $P = .04$), and a high-risk diagnosis at the time of HCT (RR, 1.32; 95% CI, 1.04-1.66; $P = .02$), older patients ($P < .01$) and recipients of partially matched URD (RR, 1.77; 95% CI, 1.21-2.58; $P < .01$) and mismatched URD or sibling donor (RR, 2.33; 95% CI, 1.70-3.20; $P < .001$). To better assess skin-only GVHD as an independent factor determining TRM at 2 years, we repeated the multivariate analysis, excluding GVHD grade as possibly confounding the effect of skin-only GVHD. This revised regression model showed that independent of GVHD grade, patients with skin-only GVHD did not have a higher risk of TRM at 2 years (RR, 1.12; $P = .24$).

In addition, 2-year TRM was higher in CMV seropositive recipients (RR, 1.28; 95% CI, 1.00-1.65; $P = .05$). Thus, these observations confirm that the adverse effect of the GVHD response at day 28 was independent and not a surrogate for other clinically important risk factors for TRM.

Discussion

The preferred and most reliable endpoint for prospective clinical trials of acute GVHD therapy has not been established. Early therapy must control GVHD and also be permissive of long-term treatment success.

We chose TRM at 2 years instead of OS as the HCT outcome to correlate with GVHD response, to reduce the confounding influence of relapse in the analysis. Our data suggest that responses at day 28 or day 56 after initiation of GVHD therapy are similarly valid as endpoints for acute GVHD trials and predict the later, more consequential outcome of 2-year TRM. Because patients who fail initial response require prompt identification and initiation of secondary therapy, we now define treatment failure by early progression or NR by day 28 as the best endpoint to define a need for new therapies. Likewise, because response at day 28 had strong predictive value for better 2-year TRM, it represents a clear and valid early endpoint to determine the effectiveness of initial GVHD therapy, which also predicts long-term success. Day 14 response is less valid because it does not correlate well with day 28 or day 56 responses or reliably predict TRM. However, CR or PR was each valuable as early measures of favorable improvement and predicted 2-year TRM. Perhaps because of our stringent definition of PR versus NR, the survival of patients with PR paralleled that of patients with CR and a substantial fraction of even the patients with early NR enjoyed 2-year survival; generally after secondary therapy for acute and later chronic GVHD.

Our results do not support the use of VGPR as the optimal endpoint for GVHD trials. However, our definition of VGPR in this retrospective analysis differs somewhat from that proposed by Martin et al[7] which incorporates a functional component to the definition, not fully addressable in a retrospective review. Similarly, a proposed acute GVHD activity index (GVHD activity index [AI])[31] also incorporates an overall performance score that may be confounded by concurrent, non-GVHD–related toxicities such as infection. Functional VGPR and GVHD AI will be difficult to assess in multicenter studies and are not testable in retrospective analyses in which performance status and oral caloric intake might be unavailable. Functional VGPR and/or acute GVHD AI could also correlate with TRM, but they need to be validated retrospectively and formally tested in prospective GVHD trials.

Although the group of patients in this analysis was heterogeneous, our results were consistent when sizeable cohorts of patients were analyzed separately, including 128 recipients of nonmyeloablative conditioning or 212 recipients of UCB. For each of these subsets, there were sufficient numbers of patients to conclude that day 28 was the best early endpoint for acute GVHD therapeutic trials. The multiple regression analysis identified no significant interactions with graft source or conditioning intensity. In this multiple regression analysis, considering all pertinent patients and transplantation variables, patients with NR at day 28 were still significantly more likely to experience TRM at 2 years than patients with a response.

In earlier trials of acute GVHD therapy, overall or organ-specific responses to therapy, TRM, or survival were used as primary endpoints, but they were most often not tested against longer term, more robust outcomes.[1-6] Reliable judgments about the value of any therapy for acute GVHD must measure both control of GVHD symptoms; be permissive of steroid withdrawal; and control the risks of opportunistic infection, chronic GHVD, and later TRM. We propose that day 28 response, including PR or CR, be incorporated as the early target, which can predict the later, and more critical outcomes for patients with acute GVHD.

Acknowledgments

research nurses whose dedicated efforts allowed prospective collection of the GVHD data.

Authorship

Contribution: All authors contributed equally to the conception, design, and interpretation of data, and the final manuscript. T.E.D.

performed the statistical analysis; and M.L.M. had primary responsibility for drafting the manuscript.

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

Correspondence: Margaret L. MacMillan, Department of Pediatrics, University of Minnesota, MMC 484, 420 Delaware St SE, Minneapolis, MN 55455; e-mail: macmi002@umn.edu.

References

1. Weisdorf D, Haake R, Blazar B, et al. Treatment of moderate/severe acute graft-versus-host disease after allogeneic bone marrow transplantation: an analysis of clinical risk features and outcome. *Blood*. 1990;75(4):1024-1030.

2. Martin PJ, Schoch G, Fisher L, et al. A retrospective analysis of therapy for acute graft-versus-host disease: initial treatment. *Blood*. 1990;76(8):1464-1472.

3. MacMillan ML, Weisdorf DJ, Wagner JE, et al. Response of 443 patients to steroids as primary therapy for acute graft-versus-host disease: comparison of grading systems. *Biol Blood Marrow Transplant*. 2002;8(7):387-394.

4. Lee SJ, Zahrieh D, Agura E, et al. Effect of upfront daclizumab when combined with steroids for the treatment of acute graft-versus-host disease: results of a randomized trial. *Blood*. 2004;104(5):1559-1564.

5. Alousi AM, Weisdorf DJ, Logan BR, et al. Etanercept, mycophenolate, denileukin, or pentostatin plus corticosteroids for acute graft-versus-host disease: a randomized phase 2 trial from the Blood and Marrow Transplant Clinical Trials Network. *Blood*. 2009;114(3):511-517.

6. Couriel D, Saliba R, Hicks K, et al. Tumor necrosis factor-alpha blockade for the treatment of acute GVHD. *Blood*. 2004;104(3):649-654.

7. Martin PJ, Bachier CR, Klingemann HG, et al. Endpoints for clinical trials testing treatment of acute graft-versus-host disease: a joint statement. *Biol Blood Marrow Transplant*. 2009;15(7):777-784.

8. Weisdorf D, Spellman S, Haagenson M, et al. Classification of HLA-matching for retrospective analysis of unrelated donor transplantation: revised definitions to predict survival. *Biol Blood Marrow Transplant*. 2008;14(7):748-758.

9. Davies SM, Wagner JE, Shu XO, et al. Unrelated donor bone marrow transplantation for children with acute leukemia. *J Clin Oncol*. 1997;15(2):557-565.

10. McGlave PB, Shu XO, Wen W, et al. Unrelated donor marrow transplantation for chronic myelogenous leukemia: 9 years' experience of the national marrow donor program. *Blood*. 2000;95(7):2219-2225.

11. Ramsay NK, Kim TH, McGlave P, et al. Total lymphoid irradiation and cyclophosphamide conditioning prior to bone marrow transplantation for patients with severe aplastic anemia. *Blood*. 1983;62(3):622-626.

12. McGlave P, Arthur D, Haake R, et al. Therapy of chronic myelogenous leukemia with allogeneic bone marrow transplantation. *J Clin Oncol*. 1987;5(7):1033-1040.

13. Kersey JH, Weisdorf D, Nesbit ME, et al. Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. *N Engl J Med*. 1987;317(8):461-467.

14. Kim TH, McGlave PB, Ramsay N, et al. Comparison of two total body irradiation regimens in allogeneic bone marrow transplantation for acute non-lymphoblastic leukemia in first remission. *Int J Radiat Oncol Biol Phys*. 1990;19(4):889-897.

15. Barker JN, Davies SM, DeFor T, Ramsay NK, Weisdorf DJ, Wagner JE. Survival after transplantation of unrelated donor umbilical cord blood is comparable to that of human leukocyte antigen-matched unrelated donor bone marrow: results of a matched-pair analysis. *Blood*. 2001;97(10):2957-2961.

16. Wagner JE, Rosenthal J, Sweetman R, et al. Successful transplantation of HLA-matched and HLA-mismatched umbilical cord blood from unrelated donors: analysis of engraftment and acute graft-versus-host disease. *Blood*. 1996;88(3):795-802.

17. Barker JN, Weisdorf DJ, DeFor TE, et al. Transplantation of 2 partially HLA-matched umbilical cord blood units to enhance engraftment in adults with hematologic malignancy. *Blood*. 2005;105(3):1343-1347.

18. Brunstein CG, Barker JN, Weisdorf DJ, et al. Umbilical cord blood transplantation after nonmyeloablative conditioning: impact on transplantation outcomes in 110 adults with hematologic disease. *Blood*. 2007;110(8):3064-3070.

19. Glucksberg H, Storb R, Fefer A, et al. Clinical manifestations of graft-versus-host disease in human recipients of marrow from HL-A-matched sibling donors. *Transplantation*. 1974;18(4):295-304.

20. Weisdorf DJ, Snover DC, Haake R, et al. Acute upper gastrointestinal graft-versus-host disease: clinical significance and response to immunosuppressive therapy. *Blood*. 1990;76(3):624-629.

21. Hings IM, Filipovich AH, Miller WJ, et al. Prednisone therapy for acute graft-versus-host disease: short- versus long-term treatment. A prospective randomized trial. *Transplantation*. 1993;56(3):577-580.

22. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus conference on acute GVHD grading. *Bone Marrow Transplant*. 1995;15(6):825-828.

23. Hings IM, Severson R, Filipovich AH, et al. Treatment of moderate and severe acute GVHD after allogeneic bone marrow transplantation. *Transplantation*. 1994;58(4):437-442.

24. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46.

25. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-481.

26. Lin DY. Non-parametric inference for cumulative incidence functions in competing risks studies. *Stat Med*. 1997;16(8):901-910.

27. Cox DR. Regression models and life tables. *J Royal Stat Soc*. 1972;(34):187-220.

28. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(6):496-509.

29. Harrel FE, Caoliff RM, Pryor DB, Lee LK, Rosati RA. Evaluating the yield of medical tests. *J Am Stat Assoc*. 1982;247(18):2543-2546.

30. Efron E, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall Inc; 1993.

31. Leisenring WM, Martin PJ, Petersdorf EW, et al. An acute graft-versus-host disease activity index to predict survival after hematopoietic cell transplantation with myeloablative conditioning regimens. *Blood*. 2006;108(2):749-755.