Response endpoints and failure-free survival after initial treatment for acute graft-*versus*-host disease

Yoshihiro Inamoto,[1] Paul J. Martin,[1,2] Barry E. Storer,[1,3] Marco Mielcarek,[1,2] Rainer F. Storb,[1,2] and Paul A. Carpenter[1,4]

[1]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle; [2]Department of Medicine, University of Washington School of Medicine, Seattle; [3]Department of Biostatistics, University of Washington School of Medicine, Seattle; and [4]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, USA

ABSTRACT

We evaluated short-term response endpoints for acute graft-*versus*-host disease treatment trials. We postulated that response endpoints should correlate with reduced symptom burden and decreased subsequent treatment failure, defined as non-relapse mortality, recurrent malignancy, or additional systemic treatment. The cohort included 303 consecutive patients who received initial systemic steroid treatment for acute graft-*versus*-host disease. Response was evaluated at day 28 after initial treatment, which in all cases preceded the onset of chronic graft-*versus*-host disease. At day 28, 36% of patients had a complete response, 26% had a very good partial response, 10% had another degree of partial response (other partial response) and 28% had no response. As expected, the symptom burden was lower in patients with very good partial response compared to those with other partial response. The frequencies of subsequent treatment failure were similar in patients with complete and very good partial responses, but lower than in patients with other partial response or no response at day 28. The frequency of second-line treatment was lower in patients with very good partial response than in those with other partial response. Risk factors associated with a lower probability of complete or very good partial response at day 28 were unrelated or human leukocyte antigen-mismatched related donor grafts and liver or gastrointestinal involvement at onset of initial treatment. Taken together, these results suggest that endpoints in acute graft-*versus*-host disease treatment trials should distinguish between very good partial response and other partial response. Our results support the use of complete or very good partial response at day 28 as an appropriate short-term primary endpoint.

Introduction

Allogeneic hematopoietic cell transplantation is frequently complicated by acute graft-*versus*-host disease (GVHD).[1,2] Three recent studies have evaluated short-term response endpoints after initial treatment of acute GVHD.[3-5] All three studies were anchored by longer-term mortality and showed that complete response (CR) and traditional partial response (PR) were associated with similar mortality rates. One of these studies examined sensitivity and specificity for predicting 6-month mortality and showed that CR/traditional PR was preferred over CR alone.[4] The authors of all three studies recommended the use of CR/traditional PR at day 28 after initial systemic treatment as an optimal short-term endpoint for clinical trials.[3-5]

Longer-term mortality, used as an anchor in previous studies, is important in evaluating short-term response endpoints for acute GVHD, but this outcome represents only one dimension of longer-term outcomes and does not account for subsequent second-line treatment, recurrent malignancy, or the onset of chronic GVHD.[3-8] The necessity of resorting to second-line treatment with additional systemic agents implies that the initial treatment did not provide the desired benefit and has been considered as treatment failure in evaluating the efficacy of treatment. Recurrent malignancy should be taken into account, since the risk of relapse could be increased if intensive immunosuppressive treatment diminishes graft-*versus*-

leukemia effects. The onset of chronic GVHD is a competing risk in studies of acute GVHD treatment. Treatment for acute GVHD is not necessarily intended to prevent chronic GVHD, and the onset of chronic GVHD interferes with any subsequent assessment of acute GVHD.

In 2009, a panel of experts proposed a new short-term response category for acute GVHD treatment studies.[9,10] The panel coined the term "very good partial response" (VGPR) to represent functionally a near complete response and proposed that VGPR might be a more appropriate endpoint than the traditional PR. Traditional PR includes both VGPR and "Other PR" that does not meet criteria for VGPR. VGPR approximates the clinical benefit of CR, since this criterion of response requires near resolution of rash and gastrointestinal abnormalities that are most bothersome to patients with acute GVHD. The panel also proposed that steroid doses at the time of response assessment could be incorporated as part of the primary endpoint in order to enhance the clinical benefit associated with the endpoint.

To investigate these concepts, we conducted a retrospective study among consecutive patients who received initial systemic steroid treatment for acute GVHD at our center. This study had three aims: (i) to characterize short-term response categories and longer-term outcomes after initial treatment of acute GVHD, (ii) to evaluate the association of short-term response categories with longer-term outcomes, and (iii) to examine the merits of incorporating a minimum percent reduction of the initial steroid dose in the response definition.

Methods

Patients

The study cohort included 303 relapse-free adult patients who received initial systemic steroid treatment for grades IIb-IV acute GVHD after a first allogeneic transplantation at the Fred Hutchinson Cancer Research Center/Seattle Cancer Care Alliance between 2000 and 2005. All patients signed consent forms allowing the use of their medical records for research related to outcomes after transplantation, and the institutional review board approved the study.

Definitions

Acute GVHD was prospectively diagnosed, staged and graded according to the established criteria.[11-13] CR was defined as the complete resolution of acute GVHD manifestations in all organs. Traditional PR was defined as an improvement in GVHD stage in at least one of the initially involved organs without complete resolution and without worsening in any other organs. VGPR was retrospectively defined when patients otherwise met the CR criteria but had at least one of the following manifestations (see *Online Supplementary Table S1* for details): (i) non-progressive stage 1 rash, not counting residual faint erythema or hyperpigmentation; (ii) resolving elevations of total serum bilirubin concentration <25% of baseline; or (iii) minimal gastrointestinal symptoms.[10] Other PR was defined as any traditional PR that did not meet criteria for VGPR. No response was defined as the same stage of GVHD in all organs or progression of GVHD in any organ.

Failure-free survival was defined by the absence of three types of treatment failure: second-line systemic treatment for acute GVHD, non-relapse mortality and recurrent malignancy during initial treatment. Second-line treatment was defined as any additional systemic treatment not used for initial treatment of acute GVHD. Second-line treatment also included an increase in steroid dose to ≥2.5 mg/kg/day because of a flare during steroid tapering. Further details of definitions are available in the *Online Supplementary Appendix*.

Prophylaxis and treatment of graft-versus-host disease

GVHD prophylaxis included a calcineurin inhibitor with either methotrexate or mycophenolate mofetil.[13] Prednisone or methylprednisolone was used for the initial treatment of acute GVHD. The initial prednisone-equivalent dose was 2 mg/kg/day in the majority of patients, in accordance with institutional standard practice, although the initial dose was 1 mg/kg/day for some patients at the attending clinicians' discretion.[13] After improvement of GVHD manifestations, steroid doses were tapered over 5 to 8 weeks according to standard institutional practice. Decisions to initiate second-line systemic treatment were made at the discretion of the attending physicians.

Statistical analysis

Cumulative incidence estimates of treatment failure, defined as the first event of recurrent malignancy, non-relapse mortality or treatment change during initial treatment, were derived, considering each event as a competing risk for the other two.[14] Onset of chronic GVHD during initial treatment was treated as a competing risk for all three types of failure. Cox regression and logistic regression models were used to identify factors predicting treatment failure and day 28 response, respectively. Sensitivity and specificity analyses were used to evaluate each response definition in predicting the absence of treatment failure at 6 months after initial treatment. Failures before response assessment were included as a non-response category.[3,4] Further details of the statistical analyses are available in the *Online Supplementary Appendix*.

Results

Patients' characteristics

The median age of patients in the study cohort was 46 years (range, 18-72 years). One hundred sixty-six (55%) patients had high-risk disease at transplantation, 248 (82%) received a mobilized blood cell graft, 85 (28%) had an HLA-mismatched donor, and 51 (17%) had reduced-intensity conditioning. The median time to initial systemic treatment for acute GVHD was 20 days (range, 5-114 days) after transplantation. At the beginning of systemic treatment, 72 patients (24%) had grades III-IV acute GVHD, and 233 (77%) were treated initially with prednisone at a dose of 2 mg/kg/day. Other characteristics are summarized in Table 1.

Outcomes after initial systemic treatment of acute graft-versus-host disease

The cumulative incidences of events after initial systemic treatment of acute GVHD are shown in Figure 1. In this analysis, only the first events of failure or chronic GVHD were considered, with the bottom area showing the incidence of first failure events and the yellow area showing the incidence of chronic GVHD as a competing risk. Among the three components of failure (treatment change, non-relapse mortality and recurrent malignancy), treatment change was the most frequent cause of failure, while recurrent malignancy and non-relapse mortality showed similar contributions to failure. The cumulative incidence of chronic GVHD as a competing risk was 0% at 28 days, 2% at 56 days and 27% at 6 months after initial treatment for acute GVHD. We used day 28 response as the short-term endpoint in further analyses, so that all patients could be evaluated before the onset of chronic GVHD.

Outcomes according to response categories at day 28 after initial treatment

Day 28 response is shown in Figure 2A. One hundred nine patients (36%) had a CR, 79 (26%) had a VGPR, 29 (10%) had an other PR and 30 (10%) had no response. Treatment failure before day 28 after initial treatment occurred in 56 (18%) patients and included treatment change (n=37; 12%), non-relapse mortality (n=15; 5%) and recurrent malignancy (n=4; 1.3%).

Landmark analyses of outcomes according to response categories at day 28 after initial treatment are shown in Figure 2B–D. These analyses included 247 patients who did not have failure before day 28. Notably, the cumulative incidence of subsequent chronic GVHD as a competing risk was similar among all response categories (overall *P*=0.14; Figure 2B). The cumulative incidence of subsequent non-relapse mortality was higher in patients with no response than in those with the other categories of response (overall *P*=0.0003; Figure 2C). The cumulative incidence of non-relapse mortality was similar among patients with CR, VGPR and other PR. In contrast, the cumulative incidence of subsequent treatment failure was higher in patients with no response or other PR compared to that in patients with CR or VGPR (overall *P*<0.0001; Figure 2D). The cumulative incidence of subsequent treatment failure was similar between patients with CR and VGPR, and between those with other PR and no response.

In multivariate analysis (Table 2), the risk of treatment failure was significantly higher for patients with other PR (HR 2.71; 95% CI, 1.51-4.86; *P*=0.0008) or no response (HR 2.60; 95% CI, 1.46-4.65; *P*=0.001) than for those with CR.

The results showed no statistically significant difference in the risk of failure for patients with other PR as compared to those with no response (HR 1.04; 95% CI, 0.52-2.07; *P*=0.91). Likewise, we found no statistically significant difference in the risk of failure for patients with VGPR as compared to those with CR (HR 1.13; 95% CI, 0.70-1.83; *P*=0.61). Other factors associated with treatment failure included older age of patients at transplantation (HR 1.40 per decade; 95% CI, 1.17-1.66; *P*=0.0002) and grades III-IV GVHD compared with grade IIb GVHD at initial treatment (HR 1.73; 95% CI, 1.11-2.70; *P*=0.02).

Outcomes at 6 months among all 303 patients according to response categories at day 28 after initial treatment are shown in Figure 3. As mentioned above, the proportion of patients with chronic GVHD as a competing risk after initial treatment of acute GVHD was similar among the four response categories. The proportions of all 6-month outcomes were closely similar between patients with CR and VGPR. The main difference between patients with VGPR and other PR was the increased proportion of subsequent treatment changes in the latter patients compared with patients with VGPR, which decreased the proportion of patients with failure-free survival without chronic GVHD among those with other PR compared to those with VGPR. The main difference between patients with other PR and no response was the higher proportion of non-relapse mortality and the lower proportions of treatment change and recurrent malignancy among patients with no response compared to those with other PR, while the proportion of patients with failure-free survival without chronic GVHD was similar between these two categories.

Correlation of response definitions with absence of treatment failure at 6 months

Table 3 summarizes the sensitivity, specificity, and positive and negative predictive values of respective response definitions correlating with absence of treatment failure at 6 months after initial treatment for acute GVHD. These analyses included all 303 patients in order to mimic the interpretation in clinical trials. The loss of sensitivity outweighed the gain of specificity with CR compared to other response definitions, suggesting that CR by itself is too stringent as an endpoint. Specificity was 10% higher for CR/VGPR than for CR/traditional PR, while sensitivity was 10% lower, indicating a balanced trade-off between specificity and sensitivity. In addition, as expected, the burden of residual GVHD manifestations at day 28 after initial treatment was lower in patients with VGPR than in those with other PR (*Online Supplementary Figure S1*).

Incorporation of steroid doses into response definition

Among 188 patients with CR/VGPR at day 28 after initial treatment, the median percentage of prednisone dose reduction was 40% (range, -79% to 100%) compared to initial doses. The extent of dose reduction was closely similar between patients treated initially with 1 mg/kg/day and those treated with 2 mg/kg/day (median, 40% *versus* 40%; *P*=0.73).

Figure 4A shows the sensitivity and specificity when a minimum percent reduction of the steroid dose was incorporated as an additional criterion of response. Specificity increased and sensitivity decreased when the minimum percent reduction of steroid dose was increased. At thresholds between ≥0% and ≥30% reduction, a small gain in specificity was offset by a balanced loss of sensitivity. At

Table 1. Patients' characteristics (N = 303).

Characteristic	N. (%)
Median age at transplant, year (range)	46 (18-72)
Sex, n. (%)	
Female	119 (39)
Male	184 (61)
Donor-patient gender combination, n. (%)	
Female to male	87 (29)
All others	216 (71)
Disease at transplant, n. (%)	
Acute myeloid leukemia	104 (34)
Myelodysplastic syndrome	89 (29)
Chronic myeloid leukemia	45 (15)
Acute lymphoblastic leukemia	20 (7)
Lymphoma	20 (7)
Chronic lymphocytic leukemia	10 (3)
Others	15 (5)
Disease risk at transplant*, n. (%)	
Low risk	137 (45)
High risk	166 (55)
Graft source, n. (%)	
Bone marrow	55 (18)
Mobilized blood cells	248 (82)
HLA and donor type, n. (%)	
HLA-matched related donor	80 (26)
HLA-matched unrelated donor	138 (45)
HLA-mismatched related donor	13 (4)
HLA-mismatched unrelated donor	72 (24)
Conditioning, n. (%)	
High intensity	252 (83)
Reduced intensity	51 (17)
GVHD prophylaxis, n. (%)	
Calcineurin inhibitor + methotrexate	213 (70)
Calcineurin inhibitor + mycophenolate mofetil	76 (25)
Other	14 (5)
GVHD grade at initial treatment, n. (%)	
IIb	231 (76)
III – IV	72 (24)
Skin stage at initial treatment, n. (%)	
Stage 0	54 (18)
Stage 1	19 (6)
Stage 2	32 (11)
Stage 3	195 (64)
Stage 4	3 (1)
Liver stage at initial treatment, n. (%)	
Stage 0	189 (62)
Stage 1	79 (26)
Stage 2	29 (10)
Stage 3	6 (2)
Stage 4	0
Gut stage at initial treatment, n. (%)	
Stage 0	124 (41)
Stage 1	139 (46)
Stage 2	23 (8)
Stage 3	16 (5)
Stage 4	1 (<1)
Organ involvement at initial treatment	
Skin only	93 (31)
Liver only	5 (2)
Gut only	14 (5)
Multiple organs	191 (63)
Initial daily steroid dose, n. (%)	
1 mg/kg	70 (23)
2 mg/kg	233 (77)

The low-risk category included chronic myeloid leukemia in chronic phase, acute leukemia and lymphoma in first remission, myelodysplastic syndrome without excess blasts and aplastic anemia. The high-risk category included all other diseases and stages.
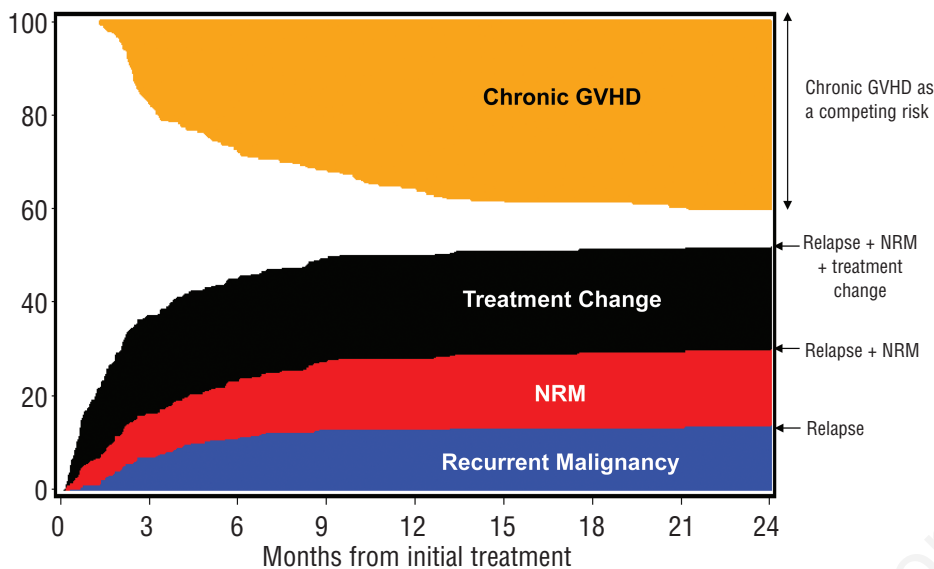
Figure 1. Categories of treatment failure after initial treatment of acute GVHD. The blue area represents failure due to recurrent malignancy. The red area represents failure due to non-relapse mortality (NRM), and the black area represents failure due to second-line systemic treatment for acute GVHD. The yellow area represents the cumulative incidence of chronic GVHD, which was treated as a competing risk.
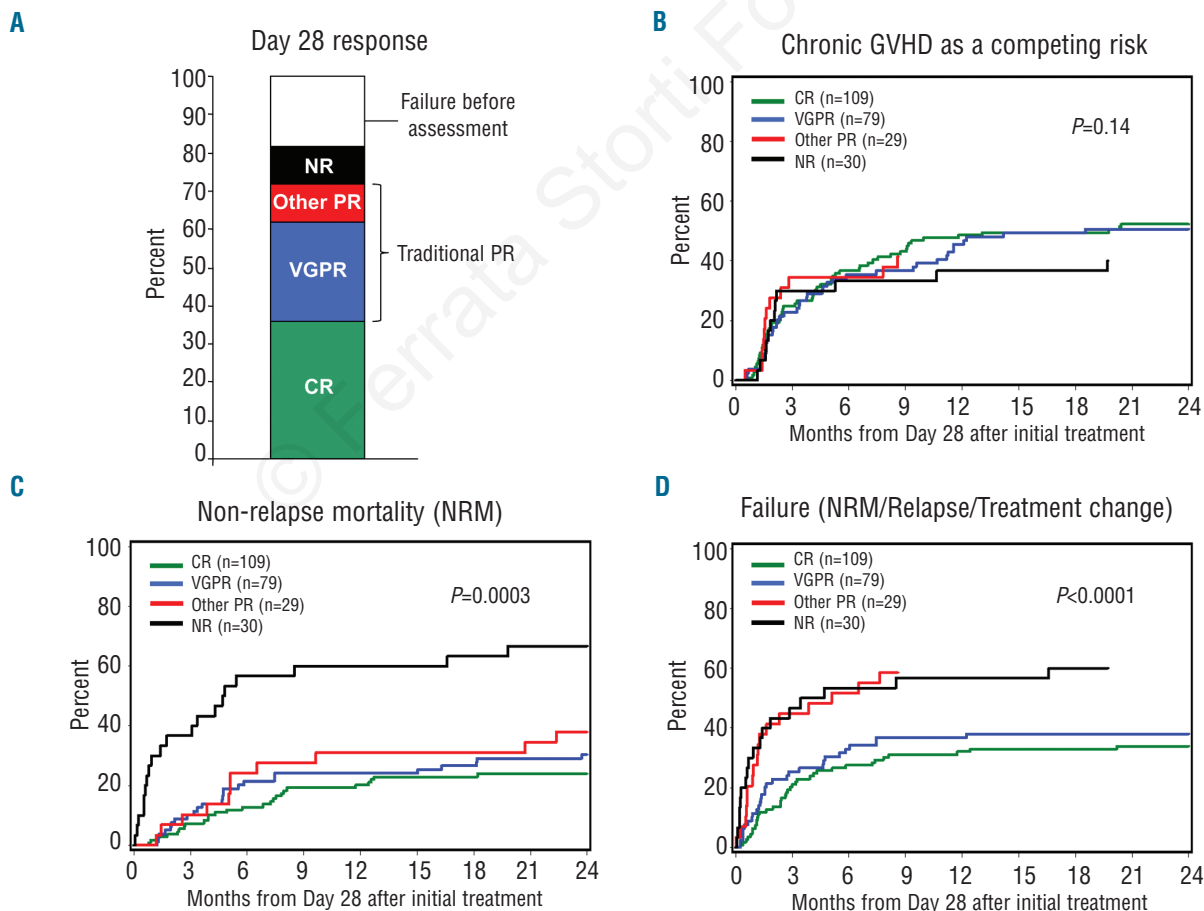

Figure 2. Day 28 response and longer-term outcomes according to response categories. (A) Response rates at days 28. Traditional partial response (PR) is subdivided into very good partial response (VGPR) and other PR. CR indicates complete response, and NR indicates no response. (B) Cumulative incidence of chronic GVHD as a competing risk according to day 28 response. (C) Non-relapse mortality (NRM) according to day 28 response. (D) Cumulative incidence of treatment failure according to day 28 response.

thresholds exceeding >30% reduction, the gain of specificity was outweighed by a large loss of sensitivity. With CR/VGPR plus ≥30% prednisone reduction as a response definition, 36 (19%) of 188 patients with CR/VGPR at day 28 were reclassified as non-responders, and specificity increased to 71%. Nonetheless, patients with ≥30% reduction and those with <30% reduction showed small differences in 6-month outcomes (Figure 4B), indicating that incorporation of steroid dose reduction in the response definition at day 28 had some, but limited value.

Factors associated with complete response/very good partial response at day 28 after initial treatment

Multivariate logistic regression models identified two factors associated with lower rates of CR/VGPR at day 28 after initial treatment: the use of an unrelated or HLA-mismatched related donor compared with the use of an HLA-matched related donor, and liver or gastrointestinal involvement at initial treatment (*Online Supplementary Table S2*). Rates of CR/VGPR at day 28 were 82% in 17 patients with neither risk factor, 74% in 139 patients with one risk factor, and 48% in 147 patients with both risk factors. Patients' age, patients' sex, disease risk, graft source, conditioning intensity, gender matching, GVHD prophylaxis, GVHD grade at initial treatment, time from transplantation to initial treatment, and initial dose of steroids were not associated with day 28 CR/VGPR.

Discussion

This study differs from previous studies in four major ways. First, we examined all three components of treatment failure, in addition to non-relapse mortality, in assessing the relationship between response categories and longer-term outcomes. In this analysis, we categorized treatment change, recurrent malignancy and non-relapse mortality as treatment failures and chronic GVHD as a competing risk. Second, we assessed VGPR and other PR separately in order to distinguish differences in the symptom burden at day 28 after initial treatment and to characterize differences in the relationship between these categories and subsequent outcomes.[10] Third, we examined the overall trade-off between sensitivity and specificity of response definitions in predicting longer-term outcomes, giving priority to specificity in order to decrease the false-positive rate (type-1 error). Fourth, we assessed the merit of incorporating a minimum percent reduction of the steroid dose as an additional criterion of response.

Our results support the use of CR/VGPR at day 28 after initial treatment as a short-term endpoint for at least four reasons. First, most patients can be evaluated at the transplant center at day 28. Moreover, none of the patients in our study developed chronic GVHD before day 28. Second, the rate of subsequent treatment failure was equivalent between patients with CR or VGPR, but was lower than that in patients with other PR or no response. Third, as expected, the symptom burden at day 28 was lower in patients with VGPR as compared to those with other PR. Fourth, CR itself was too stringent to serve as an endpoint because the loss of sensitivity outweighed the gain of specificity compared to CR/VGPR in predicting the absence of failure at 6 months.

Describing the three causes of treatment failure together with onset of chronic GVHD helped to interpret the results in this study.[15] Our results showed that 22% of patients initiated second-line treatment within 6 months and that systemic treatment change was the most frequent category of failure after initial treatment of acute GVHD. The proportions of patients who had non-relapse mortality or recurrent malignancy as reasons for treatment failure were smaller. The results shown in Figure 1 provide a useful point of comparison for future acute GVHD treatment studies.

Treatment failure after day 28 was predicted by older age of patients, grades III-IV GVHD at initial treatment and responses less than VGPR at day 28. The associations of

Table 2. Multivariate analysis of factors associated with subsequent treatment failure according to day 28 response.

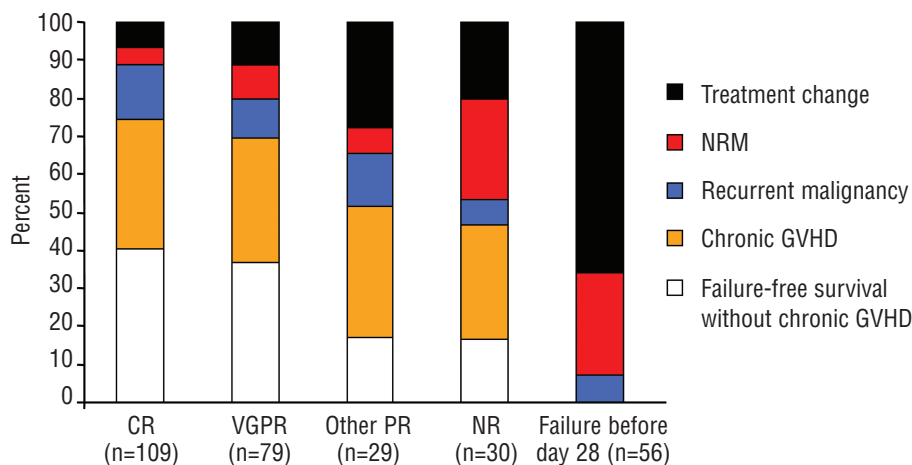| Factor	N.	Hazard ratio (95% CI)	P
Day 28 response			
Complete response	109	1.00 (reference)	
Very good partial response	79	1.13 (0.70-1.83)	0.61
Other partial response	29	2.71 (1.51-4.86)	0.0008
No response	30	2.60 (1.46-4.65)	0.001
Patients' age at transplantation (per decade)	247	1.40 (1.17-1.66)	0.0002
GVHD grade at initial treatment			
IIb	192	1.00 (reference)	
III – IV	55	1.73 (1.11-2.70)	0.02

Figure 3. Six-month outcomes according to day 28 response categories in all 303 patients. Outcomes at 6 months after initial systemic treatment were classified as treatment change, NRM, recurrent malignancy, chronic GVHD and failure-free survival without chronic GVHD. Proportions of individual outcomes at 6 months are shown according to response categories at day 28 after initial treatment. NRM: non-relapse mortality; CR: complete response; VGPR: very good partial response; PR: partial response; NR: no response.

these factors with mortality have been reported in previous studies.[3-5] Such studies showed that disease risk, HLA matching, donor relationship and liver involvement at initial treatment were associated with mortality,[3,4,16] but they were not associated with treatment failure in our study. The use of an unrelated donor or HLA-mismatched related donor and involvement of the liver or gut were associated with a lower probability of CR/VGPR at day 28 after initial treatment, consistent with the results of previous studies.[6-8] Consideration of these risk factors would help to interpret the results of other clinical studies.

The selection of day 28 after initial treatment as the time point for response assessment in our study is consistent with the findings of three previous studies that investigated endpoints for acute GVHD treatment trials.[3-5] Assessment at later time points is likely to be complicated by events that are not necessarily related to failure of acute GVHD treatment. In our analysis, we elected to treat chronic GVHD as a competing risk, since events after the onset of chronic GVHD are not necessarily related to prior acute GVHD or its treatment. These considerations make it difficult to assign outcomes as either success or failure in patients who develop chronic GVHD before the time point for response assessment. We found that patients started to develop chronic GVHD before day 56, and 27% of the patients developed chronic GVHD by 6 months after initial treatment, but no patients developed chronic GVHD before day 28. By assessing response at day 28, all patients could be evaluated before the onset of chronic GVHD.

As in previous studies, our results showed that CR, VGPR and other PR at day 28 were associated with similar subsequent non-relapse mortality.[3-5] In contrast, there were important differences in cumulative incidence of treatment failure between patients with VGPR *versus* other PR. These differences were explained by the lower proportion of treatment change in patients with VGPR than in those with other PR. The cumulative incidence of treatment failure

was similar in the VGPR group and CR group. Likewise, the cumulative incidence of treatment failure was similar in the other PR group and the no response group. In addition, by definition, patients with VGPR have a minimal symptom burden, whereas those with other PR can have a significant residual symptom burden. These comparisons emphasize the importance of distinguishing VGPR from other PR both in clinical trials and in clinical practice.

Even for patients with CR/VGPR, prolonged treatment with high-dose steroids can cause many adverse effects.[17] We, therefore, evaluated whether a minimum reduction of the initial steroid dose should be incorporated as part of the response definition at day 28 after initial treatment. In this analysis, patients with CR/VGPR and dose reduction less than the threshold were categorized as non-responders. In our study, 19% of patients with CR/VGPR at day 28 did not have a ≥30% reduction of the initial steroid dose. Reclassification of these patients as non-responders decreased the false-positive rate by 9% in predicting absence of failure at 6 months. On the other hand, the distribution of 6-month outcomes among patients with CR/VGPR did not show major differences between those with ≥30% reduction in the steroid dose compared to those with <30% reduction. This similarity contrasts with the

Table 3. Correlation of response definitions with absence of treatment failure at 6 months.

Day 28 response definition	N. of responders	Sensitivity	Specificity	PPV	NPV
CR	109	49%	80%	74%	57%
CR/VGPR	188	82%	62%	72%	75%
CR/traditional PR	217	92%	52%	70%	84%

PPV: positive predictive value; NPV: negative predictive value; CR: complete response; VGPR: very good partial response; and PR: partial response. See Methods for interpretation of the results.
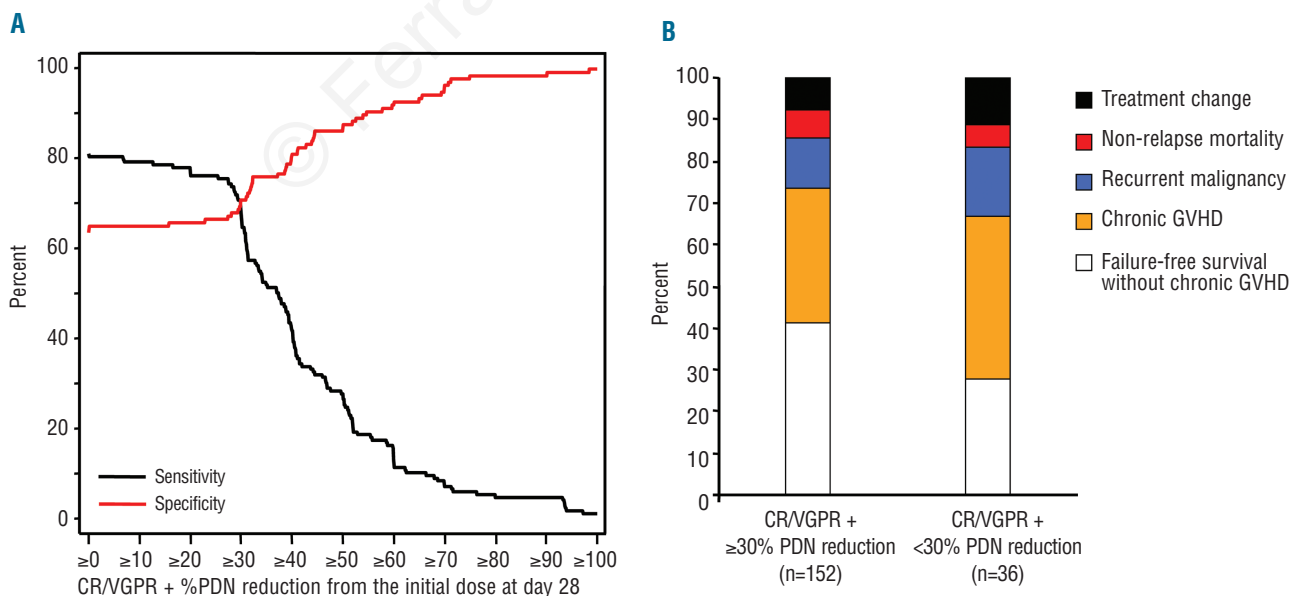

A

B

Figure 4. Effects of incorporating a minimum percent reduction of the initial steroid dose in the CR/VGPR definition of response at day 28. **(A)** Sensitivity (black line) and specificity (red line) in predicting the absence of failure at 6 months according to threshold reductions of the steroid dose. **(B)** Six-month outcomes after CR/VGPR with or without ≥30% reduction of the prednisone (PDN) dose.

large difference in outcomes between patients with VGPR compared to those with other PR. We conclude that incorporating a minimum percent reduction of the initial steroid dose has some, but limited value in the definition of response at day 28 after initial treatment for acute GVHD.

Some investigators might want to consider recurrent malignancy as a competing risk or chronic GVHD as treatment failure. The importance of distinguishing between VGPR and other PR remained the same even when different definitions were applied (*Online Supplementary Figure S2*). This was also supported by the results shown in Figure 3, since proportions of 6-month outcomes remained the same regardless of definitions of these events. Because initial treatment can diminish graft-*versus*-leukemia effects, we recommend that recurrent malignancy be considered as treatment failure in clinical trials. As shown in Figure 2B, incidence rates of chronic GVHD were similar among all response categories at day 28, indicating that effective treatment of acute GVHD does not necessarily have an influence on risk of chronic GVHD. In addition, the onset of chronic GVHD makes any subsequent events difficult to interpret from the perspective of evaluating initial treatment benefit. Thus, we recommend considering chronic GVHD as a competing risk in clinical trials.

The current study has some limitations. First, systemic treatment change and rates of steroid reduction must be carefully interpreted because both factors are controlled by providers. The value of these factors could be improved by including standardized guidelines for adding second-line treatment and for tapering steroid doses. In clinical practice, strict control of steroid tapering is difficult to enforce because patients' conditions might not allow the pre-scheduled taper. Second, the study cohort includes only adult patients who received bone marrow or growth factor-mobilized blood cells, and the results might not apply to pediatric patients or cord blood transplants. Third, 24% of our patients were treated initially with prednisone at a dose of 1 mg/kg/day, an approach that remains under investiga-

tion at our center. The aim of the current study was, however, to examine the relationship between response categories and longer-term outcomes regardless of the initial steroid dose. In fact, initial steroid dose was not associated with CR/VGPR at day 28 or with subsequent treatment failure. In addition, the conclusions held true when analyses were limited to patients treated initially with 2 mg/kg/day (results not shown). Lastly, the results were drawn from retrospective data at a single center. Further prospective and retrospective studies are warranted to determine whether findings from this study can be generalized.

This study highlights some challenges in defining the primary endpoint in acute GVHD treatment trials. Short-term response categories are poor surrogates for longer-term outcomes in patients with acute GVHD, as demonstrated by the low sensitivity and specificity in predicting 6-month outcomes in this study and in previous studies.[3-5] At the same time, longer-term outcomes cannot be used as primary endpoints, because of the difficulty of interpreting results in patients who develop chronic GVHD. These inescapable shortcomings apply to all endpoints used in acute GVHD treatment trials. Nonetheless, our analysis clearly indicates that response categories in acute GVHD treatment trials should distinguish between VGPR and other PR. Our results support the use of CR/VGPR at day 28 after initial treatment as an appropriate short-term endpoint in acute GVHD treatment trials.

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Bacigalupo A. Management of acute graft-versus-host disease. Br J Haematol. 2007; 137(2):87-98.
2. Socie G, Blazar BR. Acute graft-versus-host disease: from the bench to the bedside. Blood. 2009;114(20):4327-36.
3. MacMillan ML, DeFor TE, Weisdorf DJ. The best endpoint for acute GVHD treatment trials. Blood. 2010;115(26):5412-7.
4. Levine JE, Logan B, Wu J, Alousi AM, Ho V, Bolanos-Meade J, et al. Graft-versus-host disease treatment: predictors of survival. Biol Blood Marrow Transplant. 2010;16 (12):1693-9.
5. Saliba RM, Couriel DR, Giralt S, Rondon G, Okoroji GJ, Rashid A, et al. Prognostic value of response after upfront therapy for acute GVHD. Bone Marrow Transplant. 2012;47 (1):125-31.
6. Weisdorf D, Haake R, Blazar B, Miller W, McGlave P, Ramsay N, 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):1024-30.
7. Martin PJ, Schoch G, Fisher L, Byers V, Anasetti C, Appelbaum FR, et al. A retrospec-

tive analysis of therapy for acute graft-versus-host disease: initial treatment. Blood. 1990;76(8):1464-72.
8. MacMillan ML, Weisdorf DJ, Wagner JE, DeFor TE, Burns LJ, Ramsay NK, 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): 387-94.
9. Workshop on Clinical Trial Endpoints for Acute Graft-vs-Host Disease after Allogeneic Hematopoietic Stem Cell Transplantation. 2005 [cited 2012 September 1]; Available from: http://www.cibmtr.org/Meetings/Materials/GVHDworkshop/pages/index.aspx
10. Martin PJ, Bachier CR, Klingemann HG, McCarthy PL, Szabolcs P, Uberti JP, et al. Endpoints for clinical trials testing treatment of acute graft-versus-host disease: a joint statement. Biol Blood Marrow Transplant. 2009;15(7):777-84.
11. Przepiorka D, Weisdorf D, Martin P, Klingemann HG, Beatty P, Hows J, et al. 1994 Consensus Conference on Acute GVHD Grading. Bone Marrow Transplant. 1995;15(6):825-8.
12. Leisenring WM, Martin PJ, Petersdorf EW, Regan AE, Aboulhosn N, Stern JM, 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): 749-55.
13. Mielcarek M, Storer BE, Boeckh M, Carpenter PA, McDonald GB, Deeg HJ, et al. Initial therapy of acute graft-versus-host disease with low-dose prednisone does not compromise patient outcomes. Blood. 2009;113(13):2888-94.
14. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med. 1999;18(6):695-706.
15. Inamoto Y, Storer BE, Lee SJ, Carpenter PA, Sandmaier BM, Flowers ME, et al. Failure-free survival after second-line systemic treatment of chronic graft-versus-host disease. Blood. 2013;121(12):2340-6.
16. Robin M, Porcher R, de Castro R, Fisher G, de Latour RP, Ribaud P, et al. Initial liver involvement in acute GVHD is predictive for nonrelapse mortality. Transplantation. 2009;88(9): 1131-6.
17. Khera N, Storer B, Flowers ME, Carpenter PA, Inamoto Y, Sandmaier BM, et al. Nonmalignant late effects and compromised functional status in survivors of hematopoietic cell transplantation. J Clin Oncol. 2012; 30(1):71-7.