



Published in final edited form as:

Biol Blood Marrow Transplant. 2011 January ; 17(1): 124–132. doi:10.1016/j.bbmt.2010.06.018.

Comparison between short-term response and long-term outcomes after initial systemic treatment of chronic graft-versus-host disease

Paul J. Martin^{1,2}, Barry E. Storer^{1,3}, Paul A. Carpenter^{1,4}, Daniel R. Couriel⁵, Mary E.D. Flowers^{1,2}, Vikas Gupta⁶, Jack W. Hsu⁷, Madan Jagasia⁸, Carrie L. Kitko⁵, Richard T. Maziarz⁹, Scott D. Rowley¹⁰, Paul J. Shaughnessy¹¹, Koen van Besien¹², Daniel Weisdorf¹³, and Stephanie J. Lee^{1,2}

¹ Fred Hutchinson Cancer Research Center, Seattle, WA ² Department of Medicine, University of Washington, Seattle ³ Department of Biostatistics, University of Washington, Seattle ⁴ Department of Pediatrics, University of Washington, Seattle ⁵ University of Michigan, Ann Arbor ⁶ Princess Margaret Hospital, Toronto, ON ⁷ University of Florida, Gainesville ⁸ Vanderbilt University, Nashville, TN ⁹ Oregon Health & Science University, Portland ¹⁰ Hackensack University Medical Center, New Jersey ¹¹ Texas Transplant Institute, San Antonio ¹² University of Chicago, IL ¹³ University of Minnesota, Minneapolis

Abstract

Chronic GVHD clinical trials often use early endpoints, such as clinical response at 3 or 6 months as the primary endpoint, instead of measures of long-term treatment success, such as the ability to discontinue immunosuppressive treatment after development of immune tolerance and resolution of active disease. We evaluated the ability of defined overall and organ-specific response categories at 3 and 6 months to predict subsequent success or failure of primary treatment. The analysis included 116 patients who were evaluated at 3 months and 94 patients who were evaluated at 6 months after enrollment. Success was identified as withdrawal of systemic treatment after resolution of chronic GVHD without secondary therapy. Failure was identified as secondary systemic treatment, or death or development of bronchiolitis obliterans during primary treatment. With most definitions, response at 3 and 6 months did not show statistically significant correlation with subsequent success of primary treatment. With some definitions, the absence of response at 6 months showed statistically significant correlation with subsequent failure of primary treatment. These results suggest that early response to agents currently used for primary treatment does not necessarily predict subsequent tolerance, an important endpoint in the management of chronic GVHD. Rigorously defined clinical response is an appropriate primary endpoint for studies of chronic GVHD, but future clinical trials should provide for extended follow-up in order to ascertain late outcomes that are not necessarily predictable by evaluation of response before 6 months.

Corresponding author: Paul J. Martin, M.D., Fred Hutchinson Cancer Research Center, P.O. Box 19024, Seattle, WA 98109-1024, Phone: 206-667-4798, Fax: 206-667-5255, pmartin@fhcrc.org.

Conflict-of-interest disclosure: P.J.M. and D.W. received research funding from Roche Laboratories. All other authors declare no conflict of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

hematopoietic cell transplantation; chronic graft-versus-host disease; mycophenolate mofetil; treatment; randomized controlled clinical trial; endpoints

Introduction

A variety of endpoints have been used in clinical trials to evaluate the efficacy of treatment for chronic graft-versus-host disease (GVHD). Studies of initial treatment have evaluated response at 2–12 months, survival or nonrelapse mortality, or withdrawal of systemic treatment after resolution of the disease without secondary systemic treatment as primary endpoints. Studies of secondary treatment have typically used a measure of early outcome, such as partial or complete response, often at undefined timepoints.¹ Results from previous studies have shown that the median duration of systemic treatment for chronic GVHD is approximately 2.5–3.0 years, a period of time that is too long for typical phase II studies.² Progress in the field would be helped greatly by development of a validated shorter-term endpoint that could be used in future clinical trials to evaluate new approaches for treatment of chronic GVHD.

Data from a randomized, multicenter, blinded phase III trial testing mycophenolate mofetil (MMF) added to initial systemic treatment of chronic GVHD provided an opportunity to validate early endpoints as correlates of subsequent treatment outcomes.³ Standardized assessments of organ-specific and overall responses, and medication doses were collected every three months. Withdrawal of all systemic treatment after resolution of reversible disease manifestations without secondary systemic treatment was selected as the primary endpoint for the clinical trial, because this outcome represents cure of GVHD as the intended final goal of chronic GVHD treatment. The trial was stopped prematurely because an interim analysis suggested no improvement in outcome with the addition of MMF to initial treatment.

In the current analysis, we hypothesized that major improvement in manifestations of chronic GVHD was associated with subsequent early resolution of chronic GVHD and withdrawal of all immunosuppressive treatment without the need for secondary systemic treatment. A strong correlation between initial response and subsequent success for the primary endpoint in this study would support the use of initial response as a surrogate for subsequent success of primary treatment for chronic GVHD in future clinical trials.

Patients and Methods

Patients

Figure 1 illustrates the flow of patients in the original study.³ The current analysis is focused on the 116 patients who had responses evaluated after 3 months and on the 94 patients who had responses evaluated after 6 months. The median follow-up of patients evaluated after 3 months was 21.9 (range, 3.2–49.0) months, and the median follow-up of patients evaluated after 6 months was 22.9 (range, 6.4–49.0) months. Informed consent was documented with the use of forms approved by the Institutional Review Board of the Fred Hutchinson Cancer Research Center and the respective participating transplantation centers, in accordance with the Declaration of Helsinki.

Definitions of early responses

At enrollment and at 3-month intervals thereafter, physicians evaluated GVHD manifestations involving the skin, mouth, eyes, gastrointestinal tract, joints, genitalia and liver according to a 4-point scale of disability, similar to the scoring categories proposed by the National Institutes of Health (NIH) Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-host Disease.⁴ According to this scale, a score of 0 indicates the absence of active disease manifestations, a score of 1 indicates abnormalities associated with no more than mild disability, a score of 2 indicates moderate disability, and a score of 3 indicates severe disability. Myositis, esophageal involvement, eosinophilia and thrombocytopenia (platelet count $<100,000/\mu\text{L}$) were recorded as absent or present. Enumeration of chronic GVHD manifestations or involved sites included myositis, esophagus and eosinophilia but not thrombocytopenia. Physicians were asked to grade overall clinical severity of chronic GVHD as absent, mild, moderate or severe, without specific definitions. Changes in severity were used to infer the physician's assessment of response. Physicians were also asked to evaluate changes in overall severity from one quarterly assessment to the next, without specific definitions, but physicians were not asked directly to specify a response category. Overall NIH severity was also calculated retrospectively according to the algorithm proposed by the NIH Consensus Development Project.⁴ At baseline and at quarterly intervals thereafter, patients completed a questionnaire that included the chronic GVHD Symptom Scale, which measures the extent to which patients were bothered by manifestations of chronic GVHD.⁵

Chronic GVHD clinical manifestations assessed with the 4-point scale were considered as improved if the score at the response evaluation was less than the baseline score by at least 1 point, and manifestations were considered worse if the score increased by at least 1 point. Manifestations assessed as present or absent were scored as improved if an abnormality at baseline was absent at the response evaluation. Manifestations were scored as worse if an abnormality at the response evaluation was absent at baseline. Changes in eosinophilia and thrombocytopenia were not considered in evaluating changes between baseline and the response evaluation. Partial response was defined as improvement in at least one organ without worsening in others. Complete response was defined as a score of 0 in all organs. These criteria for partial response are less stringent than those proposed by the Response Criteria Working Group Report of the NIH Consensus Development Project.⁶ Data report forms for the trial were designed before the NIH Consensus Development Project on Criteria for Clinical Trials in Chronic GVHD.^{4,6} For this reason, we were not able to determine responses precisely according to criteria proposed in the Response Criteria Working Group Report.

Chronic GVHD symptom scores were considered improved if the summary score at 3 or 6 months was ≥ 7 points lower than the baseline score. Chronic GVHD summary symptom scores were considered worse if the score at the response evaluation was ≥ 7 points higher than the baseline score. A 7-point change in the scale represents 0.5 standard deviation of the score at baseline among participants in the MMF study, similar to results reported in the original study by Lee et al.⁵

Definitions of success and failure of primary treatment

In this study, success of primary treatment was defined as withdrawal of all systemic treatment after resolution of reversible manifestations of chronic GVHD, with no secondary systemic treatment, regardless of time from enrollment. The 2-year time limit used to define success in the original study³ was not applied in the current study. Withdrawal of treatment to improve donor chimerism or to induce an anti-tumor response after recurrent or secondary malignancy was not considered as success. Failure of primary treatment was defined as the

initiation of secondary systemic treatment or as non-relapse death or development of bronchiolitis obliterans during primary treatment.

Secondary systemic treatment included any intervention intended to control chronic GVHD through any systemic agent that was not included in the primary treatment regimen. Administration of systemic glucocorticoids to patients who were not treated initially with glucocorticoids was considered as secondary systemic treatment. Topical agents were not considered as secondary systemic treatment. An increase in the dose of prednisone and any resumption of treatment with prednisone or study drug after previous discontinuation for any reason was not considered as secondary systemic treatment. Any increase in the dose of cyclosporine or tacrolimus or resumption of treatment with cyclosporine or tacrolimus after previous discontinuation for any reason was not considered as secondary systemic treatment if the drug in question was included as part of the primary treatment regimen. A change in treatment from cyclosporine to tacrolimus or vice versa resulting from drug toxicity was not considered as secondary treatment, but any such change made because of uncontrolled chronic GVHD was considered as secondary treatment. Patients who were still receiving primary treatment when the trial was terminated were not evaluated for success or failure of primary treatment since neither endpoint was reached.

Statistical analysis

Results of the original study showed no measurable benefit when MMF was added to the initial systemic treatment for chronic GVHD.³ For this reason, the current analysis did not include consideration of whether patients received MMF or not. The cumulative incidence of success or failure of primary treatment and the associated standard error were estimated as previously described.⁷ Cox proportional hazards models were used to evaluate the association of response categories and characteristics at 3 and 6 months with subsequent success or failure of primary treatment. In these models, recurrent malignancy during primary treatment was treated as a competing risk, and follow-up was censored for patients who were still receiving primary treatment when the study ended.

Results

Testing of *a priori* response definitions

Nine *a priori* definitions of clinical responses at 3 and 6 months were initially tested for their association with subsequent success or failure of primary treatment: improvement in any organ, improvement or no change in all organs, improvement in all organs, complete response in any organ, complete response in any organ with improvement or stability in all other organs, overall partial or complete response, overall complete response, stable or improved chronic GVHD symptom score, and improved GVHD symptom score. In the original study,³ nearly all patients had success or failure of primary treatment or a competing risk within 2 years after enrollment. Among the 116 patients who were evaluated at 3 months and the 94 patients who were evaluated at 6 months after enrollment, the respective cumulative incidence rates of subsequent success of primary therapy were 25% and 28% at 2 years. Based on these results, response at 3 or 6 months according to a useful definition should predict substantially greater than a 25–28% cumulative incidence of subsequent success at 2 years. The corresponding cumulative incidence rates of subsequent failure of primary therapy were 49% and 47% at 2 years. Hence, the absence of response according to a useful definition should predict substantially greater than a 47–49% cumulative incidence of subsequent failure at 2 years.

Table 1 lists the 9 *a priori* response definitions, subdivided according to whether the response was attained at 3 months or not. The cumulative incidence of subsequent success at

2 years associated with response according to the 9 *a priori* definitions ranged from 22 to 34%, not substantially higher than the 25% overall success rate for the entire group of 116 patients without consideration of response. Response defined as a stable or improved chronic GVHD symptom score at 3 months and response defined as improvement in any organ at 3 months were associated with subsequent success, as compared to the absence of response by these definitions ($P=.01$ and $.03$, respectively). With one exception, the cumulative incidence of subsequent failure at 2 years associated with the absence of response at 3 months according to the 9 *a priori* definitions ranged from 48–72%. All 8 patients with absence of improvement in any organ at 3 months had subsequent failure of primary treatment. None of the *a priori* response definitions tested at 3 months showed a statistically significant association with subsequent failure of primary treatment.

Table 2 shows results for the 9 *a priori* response definitions tested at 6 months. The cumulative incidence of subsequent success at 2 years associated with response according to these definitions ranged from 27 to 38%, again not substantially higher than the 28% overall success rate for the entire group of 94 patients. None of the *a priori* response definitions tested at 6 months showed a statistically significant association with subsequent success of primary treatment. With one exception, the cumulative incidence of subsequent failure at 2 years associated with the absence of response at 6 months according to the 9 *a priori* definitions ranged from 47 to 73%. All 5 patients with absence of improvement in any organ at 6 months had subsequent failure of primary treatment. The absence of response defined as improvement or stability in all organs, complete response in any organ with all other organs stable or improved, and overall partial or complete response at 6 months showed statistically significant association with subsequent failure of primary treatment ($P=.01$, $.008$ and $.01$, respectively).

Differences between groups defined according to success and failure of primary treatment

Since the *a priori* definitions of overall response at 3 or 6 months generally did not predict subsequent success of primary treatment, we embarked on an exploratory search for other measures at 3 and 6 months that might predict subsequent success or failure of primary treatment. This search identified a few measures that differed between patients with subsequent treatment success and those with subsequent treatment failure. At 3 months, gastrointestinal tract scores were lower and platelet counts were higher among patients who had subsequent success than among those with subsequent failure (data not shown). At 6 months, liver scores, overall severity scores assigned by physicians, and prednisone-equivalent glucocorticoid doses were lower among patients who had subsequent success (mean \pm SD, 0.06 ± 0.10 mg/kg/day) than among those with subsequent failure (0.17 ± 0.19 mg/kg/day). Glucocorticoid doses were not available at 6 months for 4 patients, 2 with subsequent success and 2 with subsequent failure. Twenty of 21 (95%) patients in the success category had prednisone-equivalent glucocorticoid doses <0.25 mg/kg/day at 6 months, compared to 21 of 30 (70%) patients in the failure category.

In an exhaustive search, we found no statistically significant differences in change from baseline to 3 months between patients with subsequent success compared to those with subsequent failure (data not shown). We found only one statistically significant difference in change from baseline to 6 months between patients with subsequent success compared to those with subsequent failure. In the success group, all 5 patients with joint involvement showed improvement at 6 months, compared to only 1 of 6 patients with joint involvement in the failure group.

Testing of additional response definitions

Since gastrointestinal disease and thrombocytopenia at 3 months were identified as potentially important additional predictive variables, we evaluated the association between these measures and subsequent success or failure of primary treatment. In this analysis, the absence of gastrointestinal manifestations and the absence of thrombocytopenia (platelet count $<100,000/\mu\text{L}$) at 3 months did not show statistically significant association with subsequent success of primary treatment (data not shown). Likewise, the presence of thrombocytopenia at 3 months was not associated with subsequent failure of primary treatment, but the presence of gastrointestinal manifestations at 3 months showed a statistically significant association with subsequent failure of primary treatment ($P=.008$).

Prednisone-equivalent glucocorticoid dose <0.25 mg/kg/day at 6 months was added as a criterion for each of the *a priori* response definitions tested for association with subsequent outcome. Response definitions for this analysis also included a liver score of 0 and a physician severity score of 0 or 1 at 6 months with or without the additional criterion of prednisone-equivalent glucocorticoid dose <0.25 mg/kg/day. Results in Table 3 show that the cumulative incidence of subsequent success at 2 years after a response at 6 months by these definitions ranged from 28 to 45%. A liver score of 0 with or without a prednisone-equivalent glucocorticoid dose <0.25 mg/kg/day showed a statistically significant association with subsequent success ($P=.002$ and $.003$, respectively). By all other definitions, response at 6 months showed no statistically significant association with subsequent success. The cumulative incidence of subsequent failure at 2 years associated with the absence of response at 6 months for these definitions ranged from 50–84%. The absence of response by 9 definitions at 6 months showed a statistically significant association with subsequent failure, including the absence of prednisone-equivalent glucocorticoid dose <0.25 mg/kg/day, composite responses of improvement in any organ, improvement or stability in all organs, complete response in any organ, complete response in any organ with stability or improvement in all others, overall complete or partial response, and unchanged or improved GVHD symptom score, together with a prednisone-equivalent glucocorticoid dose <0.25 mg/kg/day. A physician severity score >1 at 6 months showed strong association with subsequent failure, with or without including the prednisone-equivalent glucocorticoid dose as a criterion ($P=.004$ and $.001$, respectively).

Discussion

Results of the current study did not support our hypothesis that major improvement in manifestations of chronic GVHD is associated with subsequent early resolution of chronic GVHD and withdrawal of all immunosuppressive treatment without the need for secondary systemic treatment. With most definitions, clinical response at 3 and 6 months did not predict success of primary treatment for chronic GVHD. In contrast, failure of primary treatment for chronic GVHD could be predicted according to the absence of clinical response by a variety of measures at 6 months, but not at 3 months. Most patients had clinical improvement from baseline to the response evaluations at 3 and 6 months and were able to taper systemic steroid treatment rapidly during the first 6 months regardless of subsequent failure or success of primary treatment, although systemic glucocorticoid doses were lower at 6 months for patients with subsequent success than for those with subsequent failure.

The management of chronic GVHD has 4 important goals: relief of symptoms, prevention of disability and death, and development of immunological tolerance, as indicated by resolution of chronic GVHD and permanent withdrawal of systemic treatment without exacerbation of residual disease manifestations.^{1,8} The high proportion of patients with clinical improvement during the first 3–6 months after starting treatment suggests that

systemic glucocorticoids and calcineurin inhibitors provide effective relief of symptoms. Studies during the early 1980's further indicated that in most patients, early intervention with glucocorticoids and calcineurin inhibitors prevents severe disability that would otherwise be caused by chronic GVHD.^{8,9}

Previous studies have shown correlations between certain baseline characteristics and a prolonged total duration of systemic treatment for chronic GVHD before development of tolerance. The current study analyzed outcomes of primary treatment, as opposed to the total duration of treatment, since the objective was to define a short-term response measure that could be used to indicate whether primary treatment might induce immune tolerance in future clinical trials. This approach was based on the notion that when primary treatment is successful, the total duration of treatment is likely to be shorter, and when primary treatment fails to induce tolerance, the total duration of treatment is likely to be prolonged. Withdrawal of all systemic treatment without subsequent flare of chronic GVHD manifestations serves as a robust functional test indicating that the disease has been cured, an important goal of treatment. On the other hand, one could argue that this definition of success is too stringent. In some cases, manifestations of chronic GVHD can be controlled by trivial doses of prednisone as a single treatment agent. In the current study, we did not attempt to identify such cases in order to assess whether our conclusions might change if we had used a less stringent definition of success. Likewise, we did not evaluate whether response at 3 or 6 months was associated with prevention of disability, another important goal in the management of chronic GVHD.

The general absence of correlation between initial clinical response and subsequent success of primary treatment and development of tolerance in the current study suggests that these outcomes are not functionally linked with each other in patients treated with current regimens of glucocorticoids, calcineurin inhibitors and MMF. These results suggest that agents having potent effects on relief of symptoms or prevention of disability do not necessarily have an effect on development of immune tolerance. Relief of symptoms and prevention of disability are likely to reflect the potent anti-inflammatory effect of glucocorticoids, as well as the immunosuppressive effects of other medications used to treat GVHD. Development of tolerance after hematopoietic cell transplantation presumably occurs through clonal deletion, development of anergy or non-responsiveness, or emergence of T regulatory cells,^{10,11} processes that are not necessarily facilitated by glucocorticoids and calcineurin inhibitors or by MMF in patients with chronic GVHD.

The strong association between an absence of liver involvement at 6 months and subsequent success of primary treatment is highly plausible, since previous studies have shown that hyperbilirubinemia at the onset of chronic GVHD is associated with a prolonged duration of systemic treatment.² The strong association of a physician severity score >1 at 6 months with subsequent failure of primary treatment is also highly plausible, since one would expect physicians to change systemic treatment when they perceive patients as having a high burden of persistent disease manifestations after an adequate trial of initial treatment. For future clinical trials, however, response measures that predict success of primary treatment would be far more useful than response measures that predict failure.

Our study has several limitations. First, recurrent malignancy and closure of the trial at a time when approximately one-third of patients were still receiving primary treatment limited the number of informative success and failure events and decreased statistical power to observe associations between initial response and subsequent outcomes. Point estimates for the hazard ratio of success with certain response definitions were in the range of 2.5–4.0, but confidence intervals were wide, making it difficult to draw firm conclusions. While these observations indicate a lack of robust power in these analyses, we also note that clinical

trials in chronic GVHD often enroll fewer than 100 patients, which would impose similar limitations. In retrospect, it would have been helpful if the original study had allowed longer follow-up in order to ascertain late outcomes. Collection of this data now would require reopening the study at each center. Second, since this exploratory study involved multiple comparisons, statistically significant differences at the .05 level should be interpreted with caution. Third, results might not be generally representative for patients with chronic GVHD, since less than 20% of all patients diagnosed with chronic GVHD at our center were enrolled in the study. Fourth, we were unable to assess associations between early response and subsequent survival, because the small number of deaths did not provide adequate statistical power. At least one previous study has shown a correlation between response at 6 months and subsequent survival.¹²

Responses in the current study were not categorized according to NIH criteria, because the clinical trial predated the NIH Consensus Development Project.⁶ It is noteworthy, however, that none of the response definitions ranging across an entire spectrum from the minimally stringent category of “improvement in at least one organ” to the maximally stringent category of “overall complete response” at 6 months showed a statistically significant correlation with subsequent success of primary therapy. For this reason, we believe that results with the NIH response criteria would also have shown no statistically significant correlation with subsequent success of primary therapy in our study. The absence of statistically significant correlations between early response and later outcome could reflect limitations with the numbers of patients or the types of treatment given. This negative result does not rule out the possibility that statistically significant correlations between early response and later outcome might be found in future studies with better methods for measuring response, with larger numbers of patients, with other types of treatment, or with secondary treatment.

Our results have three broad implications for future studies. First, since early response does not necessarily correlate with early development of tolerance, response should be defined in a way that clearly shows demonstrable intrinsic clinical benefit, especially in trials where response is the primary endpoint. Recommendations for measurement of response suggested by the NIH Consensus Development Project were developed as one approach toward this goal.⁶ An ideal treatment should reduce the burden of disease manifestations to an acceptable minimum and prevent disability without relying on continued treatment with glucocorticoids or other immunosuppressive agents at toxic doses.² Although disease manifestations at 6 months were similar in the success and failure groups, glucocorticoid doses were much higher among patients with subsequent failure than among those with subsequent success. Clinical responses that can be sustained with low doses of immunosuppressive treatment have greater clinical significance than those that can be sustained only with high doses of immunosuppressive treatment and their associated increased risks of opportunistic infection and long-term toxicity. For this reason, clinical trials specifying response as the primary endpoint should account for the amount of glucocorticoid treatment, or possibly the cumulative total amount of glucocorticoid treatment, as well as changes from baseline in the burden of active disease manifestations at a specified point in time after enrollment. Since steroid dose is under control of the physician, the validity of studies that incorporate steroid dose in the assessment of response would be enhanced by the use of blinded designs whenever feasible.

A second implication is that future clinical trials should provide for extended follow-up in order to ascertain late outcomes. Complete or unequivocal partial response within 6 months and a major reduction in the dose of glucocorticoids or other agents likely to cause long-term toxicity provide significant real-time clinical benefit to patients with chronic GVHD. Such responses are also likely to be associated with a greatly reduced risk of disability

associated with the disease. At the same time, however, cure of the disease represents the ultimate goal of treatment, even if it cannot serve as the primary endpoint for most studies. Extended follow-up in future studies is needed, since it cannot be assumed that response before 6 months will accurately predict subsequent cure of chronic GVHD. According to results of the previous study,³ follow-up for 2 years should suffice to assess the success or failure of primary treatment in most patients.

A third implication is that future studies should focus attention on agents that might accelerate the development of tolerance, in addition to controlling disease manifestations. The current study sponsored by the Clinical Trials Network testing extracorporeal photopheresis and sirolimus for treatment of high-risk or steroid-dependent chronic GVHD has been designed with this goal in mind. Extracorporeal photopheresis might accelerate development of tolerance by increasing the number or activity of T regulatory cells,^{13–16} while treatment with sirolimus in the absence of a calcineurin inhibitor might control T effector cells without affecting the number or function of T regulatory cells.^{17–20} The primary endpoint of the phase II component of this study will assess overall complete or partial response at 6 months after enrollment, and the primary endpoint of the phase III component will assess overall durable complete response at 2 years after enrollment. A more important difference in outcomes between arms might emerge in the development of immune tolerance as indicated by earlier withdrawal of systemic treatment after resolution of chronic GVHD manifestations.

Acknowledgments

We thank Jeanne Maffit, MaryJoy Lopez, Sheri Shanabarger, Maggie Jackson and Barbara Manion for coordinating the study; Jane Jocom for nursing support; Peggy Adams Myers for administrative assistance; Dr. Katherine Guthrie for statistical support to the DSMB; Drs. Donna Neuberger, Joachim Deeg, Andrew Gilman, John Zaia and Steve Pavletic and Ms. Susan Stewart for service as members of the DSMB; Caroline McKallor and Geoff Hirschi for assistance with data management; Sheree Miller and staff at the University of Washington and Daniel McMannis and staff at Costco for pharmacy services. We thank the coordinating liaisons and study nurses at each of the participating centers for assistance with implementing the study at each participating site. We thank Stuart Tenney for assistance in preparing the manuscript and Dr. Yoshihiro Inamoto for reviewing the manuscript. We especially thank the patients who agreed to participate in the study.

Participating centers included the Fred Hutchinson Cancer Research Center, University of Minnesota, Hackensack University Medical Center, University of Florida, Baylor University Medical Center at Dallas, Stanford University, University of Nebraska, Texas Transplant Institute, Vanderbilt University, University of Chicago, City of Hope Medical Center, Oregon Health and Science University, M.D. Anderson Cancer Center, Princess Margaret Hospital, and the University of Michigan.

This research was supported by grant CA98906 from the National Cancer Institute, Department of Health and Human Services, by a grant from Roche Laboratories, represented by Dr. Kristine Golebski, and by a grant from the Office of Naval Research, administered through the National Marrow Donor Program. The double-blind design of this study would not have been feasible without the study drug generously supplied by Roche Laboratories.

References

1. Lee, SJ. Chronic graft-versus-host disease. In: Atkinson, K.; Champlin, R.; Ritz, J.; Fibbe, WE.; Ljungman, P.; Brenner, MK., editors. *Clinical Bone Marrow and Blood Stem Cell Transplantation*. Cambridge, UK: Cambridge University Press; 2004. p. 1133-1157.
2. Stewart BL, Storer B, Storek J, et al. Duration of immunosuppressive treatment for chronic graft-versus-host disease. *Blood* 2004;104(12):3501–3506. [PubMed: 15292060]
3. Martin PJ, Storer BE, Rowley SD, et al. Evaluation of mycophenolate mofetil for initial treatment of chronic graft-versus-host disease. *Blood* 2009;113(21):5074–5082. [PubMed: 19270260]
4. Filipovich AH, Weisdorf D, Pavletic S, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and Staging Working Group report. *Biol Blood Marrow Transplant* 2005;11(12):945–956. [PubMed: 16338616]

5. Lee SJ, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant* 2002;8(8):444–452. [PubMed: 12234170]
6. Pavletic S, Martin P, Lee SJ, et al. Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: IV. Response Criteria Working Group report. *Biol Blood Marrow Transplant* 2006;12(3):252–266. [PubMed: 16503494]
7. 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. [PubMed: 10204198]
8. Sullivan KM, Shulman HM, Storb R, et al. Chronic graft-versus-host disease in 52 patients: adverse natural course and successful treatment with combination immunosuppression. *Blood* 1981;57(2):267–276. [PubMed: 7004534]
9. Sullivan KM, Witherspoon RP, Storb R, et al. Alternating-day cyclosporine and prednisone for treatment of high-risk chronic graft-versus-host disease. *Blood* 1988;72(2):555–561. [PubMed: 3042042]
10. Welniak LA, Blazar BR, Murphy WJ. Immunobiology of allogeneic hematopoietic stem cell transplantation (Review). *Annu Rev Immunol* 2007;25:139–170. [PubMed: 17129175]
11. Nguyen VH, Zeiser R, Negrin RS. Role of naturally arising regulatory T cells in hematopoietic cell transplantation (Review). *Biol Blood Marrow Transplant* 2006;12(10):995–1009. [PubMed: 17084366]
12. Arora M, Burns LJ, Davies SM, et al. Chronic graft-versus-host disease: A prospective cohort study. *Biol Blood Marrow Transplant* 2003;9:38–45. [PubMed: 12533740]
13. Greinix HT, Socie G, Bacigalupo A, et al. Assessing the potential role of photopheresis in hematopoietic stem cell transplant. *Bone Marrow Transplant* 2006;38(4):265–273. [PubMed: 16883310]
14. Peritt D. Potential mechanisms of photopheresis in hematopoietic stem cell transplantation (Review). *Biol Blood Marrow Transplant* 2006;12(1 Suppl 2):7–12. [PubMed: 16399596]
15. Biagi E, Di B I, Leoni V, et al. Extracorporeal photochemotherapy is accompanied by increasing levels of circulating CD4+CD25+GITR+Foxp3+CD62L+ functional regulatory T-cells in patients with graft-versus-host disease. *Transplantation* 2007;84(1):31–39. [PubMed: 17627234]
16. Gatz E, Rogers CE, Clouthier SG, et al. Extracorporeal photopheresis reverses experimental graft-versus-host disease through regulatory T cells. *Blood* 2008;112(4):1515–1521. [PubMed: 18411417]
17. Zeiser R, Nguyen VH, Beilhack A, et al. Inhibition of CD4+CD25+ regulatory T-cell function by calcineurin-dependent interleukin-2 production. *Blood* 2006;108(1):390–399. [PubMed: 16522809]
18. Zorn E, Nelson EA, Mohseni M, et al. IL-2 regulates FOXP3 expression in human CD4+CD25+ regulatory T cells through a STAT-dependent mechanism and induces the expansion of these cells in vivo. *Blood* 2006;108(5):1571–1579. [PubMed: 16645171]
19. Coenen JJ, Koenen HJ, van Rijssen E, et al. Rapamycin, not cyclosporine, permits thymic generation and peripheral preservation of CD4+ CD25+ FoxP3+ T cells. *Bone Marrow Transplant* 2007;39(9):537–545. [PubMed: 17351648]
20. Zeiser R, Leveson-Gower DB, Zambricki EA, et al. Differential impact of mammalian target of rapamycin inhibition on CD4+CD25+Foxp3+ regulatory T cells compared with conventional CD4+ T cells. *Blood* 2008;111(1):453–462. [PubMed: 17967941]

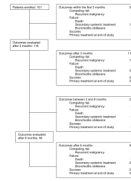


Figure 1. Flow diagram

The figure shows outcomes of primary treatment before and after 3 and 6 months from enrollment in the study.

Table 1

A priori response definitions tested at 3 months for correlation with subsequent success or failure of primary therapy (N=116)

Response Definition	3-month Response	N	Cumulative Incidence Success [†]	P [‡]	Cumulative Incidence Failure [‡]	P [‡]
Improved in any organ	Yes	108	26 ± 5	.03	48 ± 6	.23
	No	8	0		100	
Improved or stable in all organs	Yes	84	25 ± 6	.23	43 ± 7	.16
	No	32	28 ± 9		60 ± 10	
Improved in all organs	Yes	40	22 ± 8	.94	29 ± 9	.08
	No	76	26 ± 6		58 ± 7	
Complete response (CR) in any organ	Yes	102	27 ± 5	.14	49 ± 6	.15
	No	14	9 ± 8		48 ± 20	
CR in any organ, all others stable or improved	Yes	74	27 ± 6	.97	44 ± 7	.77
	No	42	21 ± 7		58 ± 10	
Overall partial or complete response	Yes	79	25 ± 6	.63	44 ± 7	.56
	No	37	24 ± 8		63 ± 10	
Overall complete response	Yes	27	26 ± 10	.50	30 ± 10	.27
	No	89	25 ± 6		54 ± 7	
GVHD symptom score unchanged or improved*	Yes	92	30 ± 6	.01	44 ± 6	.08
	No	16	0		72 ± 14	
GVHD symptom score improved*	Yes	38	34 ± 9	.90	45 ± 10	.48
	No	70	22 ± 6		48 ± 7	

* Total of "yes" and "no" is less than 116 because of missing data.

[†] at 2 years ± standard error

[‡] derived from Cox model

Table 2

A priori response definitions tested at 6 months for correlation with subsequent success or failure of primary therapy (N=94)

Response Definition	6-month Response		N	Cumulative Incidence Success [†]	P [‡]	Cumulative Incidence Failure [†]	P [‡]
Improved in any organ	Yes		89	30 ± 6	.11	43 ± 6	.19
	No		5	0		100	
Improved or stable in all organs	Yes		61	29 ± 7	.87	37 ± 8	.01
	No		33	26 ± 9		63 ± 10	
Improved in all organs	Yes		28	27 ± 10	.78	39 ± 12	.12
	No		66	28 ± 7		51 ± 7	
Complete response (CR) in any organ	Yes		80	30 ± 6	.55	41 ± 7	.22
	No		14	16 ± 11		73 ± 13	
CR in any organ, all others stable or improved	Yes		58	30 ± 7	.98	36 ± 8	.008
	No		36	25 ± 8		64 ± 10	
Overall partial or complete response	Yes		61	29 ± 7	.87	37 ± 8	.01
	No		33	26 ± 9		63 ± 10	
Overall complete response	Yes		22	33 ± 11	.36	42 ± 13	.22
	No		72	26 ± 6		49 ± 7	
GVHD symptom score unchanged or improved*	Yes		75	28 ± 6	.77	43 ± 7	.07
	No		11	21 ± 13		68 ± 15	
GVHD symptom score improved*	Yes		31	38 ± 10	.54	45 ± 11	.79
	No		55	21 ± 6		47 ± 8	

* Total of "yes" and "no" is less than 94 because of missing data.

[†] at 2 years ± standard error

[‡] derived from Cox model

Table 3

Modified response definitions tested at 6 months for correlation with subsequent success or failure of primary therapy (N=94)

Modified Response Definition	6-month Response*		N	Cumulative Incidence Success [†]	P [‡]	Cumulative Incidence Failure [†]	P [‡]
Prednisone dose <0.25 mg/kg/day	Yes		74	30 ± 6	.25	39 ± 7	.04
	No		15	9 ± 9		80 ± 12	
Improved in any organ [¶]	Yes		71	31 ± 6	.17	37 ± 7	.009
	No		18	7 ± 7		84 ± 10	
Improved or stable in all organs [¶]	Yes		50	30 ± 8	.84	33 ± 8	.009
	No		39	24 ± 8		63 ± 9	
Improved in all organs [¶]	Yes		23	28 ± 11	.55	38 ± 13	.26
	No		66	26 ± 6		50 ± 7	
Complete response (CR) in any organ [¶]	Yes		62	32 ± 7	.41	34 ± 7	.008
	No		27	15 ± 8		74 ± 10	
CR in any organ, all others stable or improved [¶]	Yes		47	30 ± 8	.70	31 ± 8	.006
	No		42	23 ± 8		64 ± 9	
Overall partial or complete response [¶]	Yes		50	30 ± 8	.84	33 ± 8	.009
	No		39	24 ± 8		63 ± 9	
Overall complete response [¶]	Yes		19	34 ± 12	.27	38 ± 14	.32
	No		70	25 ± 6		50 ± 7	
GVHD symptom score unchanged or improved [¶]	Yes		64	31 ± 7	.75	37 ± 7	.02
	No		22	18 ± 9		70 ± 11	
GVHD symptom score improved [¶]	Yes		26	45 ± 11	.33	36 ± 11	.37
	No		60	20 ± 6		50 ± 8	
Liver score 0	Yes		67	39 ± 7	.003	43 ± 8	.26
	No		27	4 ± 4		56 ± 11	
Liver score 0 [¶]	Yes		53	41 ± 8	.002	36 ± 8	.16

Modified Response Definition	6-month Response*	N	Cumulative Incidence Success [†]	P [‡]	Cumulative Incidence Failure [†]	P [‡]
	No	36	6 ± 4		61 ± 10	
Physician score 0 or 1	Yes	79	28 ± 6	.83	40 ± 7	.001
	No	11	18 ± 12		82 ± 12	
Physician score 0 or 1 [¶]	Yes	65	29 ± 7	.71	34 ± 7	.004
	No	21	18 ± 9		76 ± 11	

* Total of "yes" and "no" in most categories is less than 94 because of missing data.

[†] at 2 years ± standard error

[‡] derived from Cox model

[¶] with prednisone dose <0.25 mg/kg/day