This PDF contains briefing information submitted to FDA by members of the Clinical Trial Design Working Group in preparation for a formal pre-IND meeting on January 28, 2015 to discuss possible endpoints in clinical trials to evaluate products for treatment of chronic GVHD. The meeting was not intended to discuss any specific products.

The briefing document includes a cover letter, background information, an extensive presentation of data relevant to 5 proposed endpoints for chronic GVHD treatment trials, including 21 tables, 8 figures, references, 3 supplementary tables, and 15 appendices. The PDF also includes preliminary queries from FDA, a response to FDA and a copy of the official meeting minutes.

Hyperlinks are provided for publications included as appendix items. Appendices VIII, IX and XV have not been published and are included in the PDF. Although Appendix XV was submitted in the briefing package, it was not discussed with FDA and does not represent consensus views of the Clinical Trial Design Working Group.

Bookmarks have been included with the PDF to facilitate navigation to items of interest. The following Table of Contents is numbered according page in the PDF. Page numbers in the original submission to FDA have not been removed.

A copy of this PDF with any updates is available at http://www.asbmt.org/?page=PIND124475.

Item **PDF Page**

Paul J. Martin, MD
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N, D2-100
P.O. Box 19204
Seattle, WA 98109-1024
Phone: (206)-667-4798
FAX: (206)-667-5155
e-mail: pmartin@fhcrc.org

December 29, 2014

Ann T. Farrell, MD
Food and Drug Administration
Center for Drug Evaluation and Research
White Oak Building 22, Room: 2309
10903 New Hampshire Avenue
Silver Spring, Maryland 20993

Re: PIND 124475

Dear Dr. Farrell,

We appreciate the opportunity to meet with you and your colleagues to discuss endpoints that could be used in the development path for products related to treatment of chronic GVHD. We would benefit from a better understanding of the extent to which these endpoints could be viewed as indicators of clinical benefit for purposes of regulatory review at each phase of the development path.

As you will see in the attached briefing package, we propose 5 endpoints for consideration: 1) failure-free survival, 2) survival without progressive impairment, 3) complete or partial response, 4) patient-reported outcomes, and 5) an aggregate scale incorporating several different types of measures, similar to scales used for regulatory review of autoimmune diseases. We have not included a discussion of survival to resolution of chronic GVHD and withdrawal of all systemic treatment, survival without recurrent malignancy or overall survival, since these endpoints clearly indicate clinical benefit.

We would like to know whether you and your colleagues agree that these 5 endpoints could be used as secondary endpoints in any trial. Given the data in the briefing package, we would also like to know whether each of these might be acceptable as the primary endpoint in a “pivotal” trial, and if so, which ones would be preferable from a regulatory perspective. If data in the briefing package are not sufficient to draw a conclusion, we would appreciate any suggestions for additional data that would help address the question.

In Appendix XV of the briefing package, we provide some scenarios indicating how we envision that the proposed endpoints could be used in several development paths. It would be helpful to know whether these approaches would be considered acceptable or not.

Please let us know if any additional information in advance of the meeting would assist your review. We look forward to a productive discussion with you and your colleagues on January 28, 2014.

Sincerely yours,

Paul J. Martin, MD
Member, Fred Hutchinson Cancer Research Center
Professor of Medicine, University of Washington

Type B Meeting, January 28, 2015

Clinical Trial Endpoints in the Development of Products for Treatment of Chronic Graft-versus-host Disease

Acknowledgments

We are especially grateful to Dr. Barry Storer and Shawn Chai for assistance with statistical analysis. We also acknowledge Dr. Scott Emerson for suggesting survival without progressive impairment as a clinically meaningful endpoint in clinical trials testing products for treatment of chronic graft-versus-host disease.

Conflicts of Interest

Paul Martin: Scientific Advisory Board Meeting for Pharmacyclics, Inc.

Corey Cutler: Pharmacyclics, Inc., Onyx, Inc., Immucor, Inc., Fate Therapeutics

John Koreth: Research funding: Millennium, Otsuka, Prometheus Labs, Inc.
Advisory Board: Takeda
CME: OptumHealth

Stephanie Lee: Scientific Advisory Board Meeting for Bristol Myers Squibb

Table of Contents

Purpose of the Meeting

The purpose and objectives of this meeting are to identify endpoints that measure clinical benefit across different trial phases in the development of products indicated for treatment of chronic GVHD. No clinical development path has been mapped for indications related to treatment of chronic GVHD, and no products have been approved for such indications. Much of the current difficulty originates from the lack of a validated global clinical scoring system that could be used to measure response in studies intended for regulatory review. This pre-IND review provides an opportunity for a regulatory assessment of 5 potential clinical trial endpoints currently under active discussion in the academic community. These include failure-free survival, survival without progressive impairment, clinical response, patient-reported outcomes, and an aggregate measure incorporating provider and patient assessments. Chronic GVHD investigators would benefit from better understanding of the extent to which these endpoints could be viewed as indicators of clinical benefit for purposes of regulatory review at each phase of the development path.

Background

Development of more effective treatments for chronic GVHD is an urgent unmet clinical need. Pharmaceutical interest in the problem of chronic GVHD has been hampered by the lack of defined pathways for clinical development and regulatory approval of products intended for treatment of chronic GVHD. Regulatory applications are most likely to come as new indications for approved products, but in certain cases, they could also come as new products for the specific indication of chronic GVHD. In our experience, the number of pharmaceutical companies interested in developing products for GVHD has increased markedly since the 2005 NIH Consensus Conference. This observation and the upcoming publication of the 2014 NIH Consensus Conference Clinical Trials Working Group Report highlight the importance of efforts to clarify the merits of each potential endpoint for the field.

The number of patients available for enrollment in clinical trials evaluating products for treatment of chronic GVHD is limited. In the U.S., approximately 8,000 allogeneic hematopoietic cells transplants are now done each year.[1] Among these, at least 35% would be expected to develop chronic GVHD requiring systemic treatment,[2] such that the total incidence is approximately 3,000 per year. Rates of death and recurrent malignancy during treatment after the onset of chronic GVHD have been reported together with rates of withdrawal of immunosuppression after resolution of chronic GVHD.[3] (See Appendix IV.) Based on these rates, the total prevalence in the U.S. is estimated at less than 10,000.

To date, results from 6 randomized trials for initial treatment of chronic GVHD have been published (Supplementary Table 1),[4-9] but none of these studies demonstrated superiority of the investigational arm. The design of controlled second-line treatment studies is hampered by the lack of a standard treatment regimen. Only 1 randomized trial for second-line treatment of chronic GVHD has been published.[10,11]

Uncontrolled single-arm studies of second-line treatment typically show overall response rates of 30 – 70%.[12] In many studies, response criteria are poorly defined, and results are interpreted under the

premise that no response would have occurred in the absence of the investigational treatment. This premise might not hold true, especially if the prior trajectory of the disease and the effects of other elements in the treatment regimen are taken into account. These include changes in the doses of concomitant systemic medications and addition of topically active agents implemented at the same time when the investigational treatment was started or at any time after enrollment but before the assessment of response. These factors and variation in selection criteria have made it difficult to establish adequate benchmark response rates for statistical analysis in single arm studies.

Goals of Treatment for Chronic GVHD

Treatment of chronic GVHD is intended to produce a sustained benefit by reducing symptom burden, controlling objective manifestations of disease activity and preventing damage and impairment, without causing disproportionate harms related to the treatment itself. The goals of treatment for chronic GVHD are highlighted and brought into focus by considering the adverse outcomes reported in a small group of 13 patients with "extensive" chronic GVHD that was not treated (Appendix I).[13] Only 2 of these patients survived for more than 2 years with Karnofsky scores ≥70. Progressive oral and ocular sicca syndromes, pulmonary and hepatic insufficiency, scleroderma-like skin disease and contractures caused considerable morbidity. Seven of the 13 patients died within the first 2 years after diagnosis. Four of the 6 surviving beyond 2 years and 2 patients surviving for less than 2 years had disabling contractures. None of the 6 patients with contractures had evidence of spontaneous improvement.

Management of chronic GVHD has relied on corticosteroids as the mainstay of treatment for more than 3 decades, although the exact administration regimen varies (Appendices II and III).[14,15] Systemic treatment typically begins with prednisone at 0.5 to 1mg/kg/day, with or without cyclosporine, tacrolimus or sirolimus. Prolonged treatment with prednisone at high doses causes many adverse effects, making it necessary to taper the dose as soon as GVHD improves. In a recent prospective study, the average dose of prednisone was tapered to 0.20 – 0.25 mg/kg/day or 0.4 – 0.5 mg/kg every other day within 3 months after starting systemic treatment.[8] Manifestations of chronic GVHD can reappear or worsen when the intensity of immunosuppressive treatment is closely calibrated to the minimum dose needed to control GVHD (Figure 1). Therefore, clinical trials must allow some flexibility in the management of steroid doses, and reescalation should be allowed without designating such events as treatment failure, unless a new systemic medication is added.

When immunological tolerance develops in patients with chronic GVHD, systemic treatment can be withdrawn without risk of recurrent chronic GVHD. Low-level immunological activity may persist but is not sufficient to cause clinical manifestations of the disease. Clinical tolerance emerges in approximately 50% of patients within 7 years after starting systemic treatment, as indicated by permanent withdrawal of systemic treatment without subsequent recurrence of disease activity or exacerbation of any residual damage. Approximately 10% of patients require continued systemic treatment for an indefinite period beyond 7 years, and the remaining 40% have recurrent malignancy or die within 7 years after diagnosis (Appendix IV).[3]

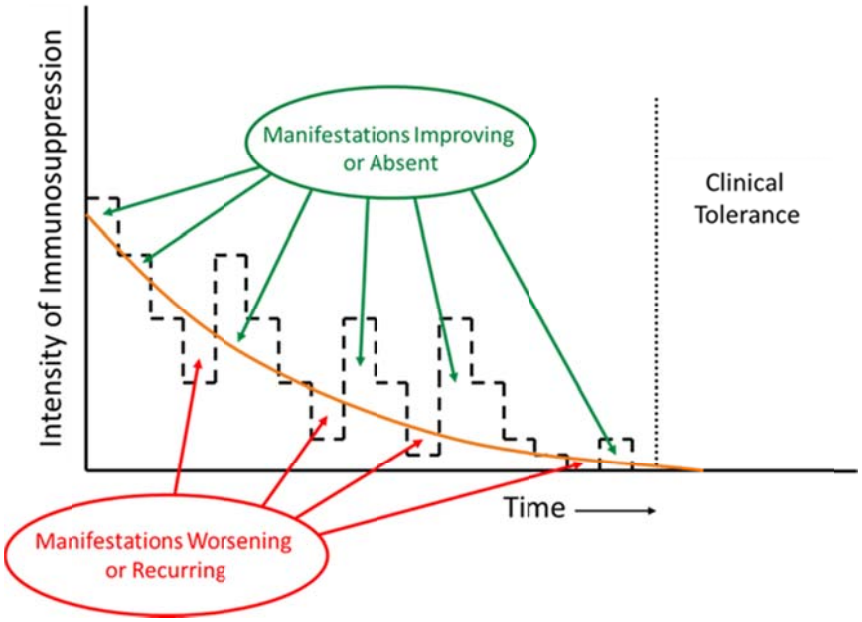


Figure 1. Appropriate management of chronic GVHD requires continuous recalibration of immunosuppressive treatment in order to avoid over- or under-treatment. The intensity of treatment required to control the disease decreases across time. Manifestations of chronic GVHD improve or are absent when the intensity of treatment (- - -) is above the threshold shown as the orange curve, and they worsen or recur when the intensity of treatment is below the threshold. The slope of the threshold varies among patients and can be determined only by serial attempts to decrease the intensity of treatment. Clinical tolerance is defined by the ability to withdraw all systemic treatment without recurrence of chronic GVHD.

It is not known whether currently available immunosuppressive products accelerate or retard the development of tolerance. Even if they do not, they provide clinical benefit by controlling disease activity and preventing impairment until tolerance develops. In this context, new products for treatment of chronic GVHD could increase clinical benefit if they are more effective than currently available treatments without causing a disproportionate burden of side effects, or if they are as effective as currently available treatment but cause a lesser burden of side effects.

Endpoints in Clinical Trials for Treatment of Chronic GVHD—2005

The 2005 NIH Consensus Conference on Chronic GVHD Clinical Trials Working Group Report addressed a variety of technical and quality considerations in the design and conduct of clinical trials testing products for treatment of chronic GVHD (Appendix V).[16] Potential short-term primary and secondary endpoints discussed in the report included GVHD response and patient reported outcomes. The report noted that scales for measurement of global response have not yet been validated and that few sensitive instruments are available for measuring patient-reported outcomes. As summarized in Table 1, GVHD response was considered most appropriate as a primary endpoint in phase II studies and possibly in selected phase III studies, while patient-reported outcomes were considered appropriate as secondary endpoints. Complete response and development of clinical tolerance were considered most appropriate

as primary endpoints in phase III studies, while non-relapse mortality, survival without recurrent malignancy and overall survival were considered appropriate as secondary endpoints.

Table 1. Endpoint recommendations in the 2005 Working Group Report

Time horizon	Primary endpoint	Secondary endpoints
Short	GVHD response	Patient-reported outcomes
Long	Complete response	Non-relapse mortality
	Clinical tolerance*	Survival without recurrent malignancy
		Overall survival

*permanent discontinuation of all systemic treatment without subsequent recurrence of disease activity or exacerbation of residual damage

Provisional Summary Assessment of Proposed Endpoints—2014

The proposed endpoints of failure-free survival and survival without progressive impairment respectively measure benefit indirectly or directly as the absence of new harm caused by the disease, whereas response measures benefit as improvement in manifestations of the disease. Patient-reported outcomes capture patient perception of GVHD symptoms and impact. Aggregate outcomes capture clinician assessments, patient-reported outcomes and laboratory or functional measures in a single global scale. All 5 endpoints represent relatively short-term outcomes as compared to the typical 2 to 5 year duration of treatment needed before the disease resolves with currently available regimens. Therefore, an important issue is the extent to which these short-term endpoints indicate longer-term outcomes.

Table 2. Strengths and weaknesses of five proposed endpoints in chronic GVHD therapy trials

Proposed endpoint	Definition	Statistical considerations	Strengths	Weaknesses
Failure free survival	Survival without new systemic treatment, death or recurrent malignancy	• Time-to-event, or • Comparison of proportions with failure-free survival at a specific time point	• Benchmarks available for 1st and 2nd-line treatment • Correlates with overall improvement reported by providers and patients • Correlates with ability to discontinue systemic treatment	• Indirect measure of failure • Improvement is not measured • New treatment decisions are subject to bias and inconsistency
Survival without progressive impairment	Survival without an enduring chronic GVHD-related effect that threatens or compromises physical well-being or function in ways that cannot be easily reversed	• Time-to-event, or • Comparison of proportions surviving without progressive impairment at a specific time point	• Failure directly measured • Correlates with overall improvement reported by providers and patients	• Improvement is not measured • Impairment not yet fully defined • Some impairment measures might not be entirely specific for chronic GVHD
GVHD Response	Complete plus partial response based on clinician-reported measures	• Comparison of proportions with treatment response at a specific time point	• Direct measure of success • Lengthy follow-up not needed • Easily applied	• Scales not fully qualified
Patient-reported outcomes	Self-reported patient information on symptoms and multi-dimensional quality of life	• Comparison of proportions with clinically meaningful improvement at a specific time point • Comparison of distributions between study arms	• Captures the patient perspective • Lengthy follow-up not needed • Easily applied	• Subject to respondent biases • Missing data difficult to control • Claims limited to PROs
Aggregate scale	Selected measures from provider and patient	• Comparison of proportions with clinically meaningful improvement at a specific time point • Comparison of distributions between study arms	• Aggregates data from multiple perspectives	• Scale not developed or qualified

The following sections address each of these endpoints in turn, with specific questions as outlined in the meeting request.

Failure-free Survival (FFS)

For this endpoint, "failure" has been defined as death, recurrent or progressive malignancy, or the initiation of new systemic treatment for chronic GVHD.[17] Increased dosing of existing treatment is not considered as failure. The premise underpinning this endpoint is that chronic GVHD was adequately controlled in cases where no new systemic treatment was given and that GVHD was not adequately controlled in cases where new systemic treatment was given.

This section poses 4 questions related to the use of "failure-free survival" as a primary endpoint in chronic GVHD treatment trials.

a. To what extent could failure-free survival at 12 months be considered as an indicator of clinical benefit in early phase trials of initial systemic treatment for chronic GVHD?

b. To what extent could failure-free survival at 6 months be considered as an indicator of clinical benefit in early phase trials of second-line systemic treatment for chronic GVHD?

c. Should the absence of recurrent or progressive malignancy be included as a component in the definition of failure-free survival?

d. Should steroid doses below a predefined threshold at 12 months after initial treatment or at 6 months after second-line treatment be included as an additional criterion of failure-free survival?

The answers to the last 2 questions in this section could influence the answers to the first two questions. Therefore, questions related to this endpoint are highly interrelated and should be considered in aggregate as well as individually. A general discussion of this endpoint follows the data presentation.

a. To what extent could failure-free survival at 12 months be considered as an indicator of clinical benefit in early phase trials of initial systemic treatment for chronic GVHD?

Correlation of failure-free survival at 12 months with subsequent outcomes

The following study illustrates how FFS serves as an intermediate endpoint predicting survival and cure of chronic GVHD. A large landmark analysis tested whether addition of a new systemic treatment by 12 months was associated with subsequent survival.[17] (See Appendix VI for full details.) Patients were analyzed in 3 groups based on events during the first 12 months after starting systemic treatment for chronic GVHD: those who had recurrent malignancy, those without recurrent malignancy who received second-line systemic treatment for chronic GVHD, and those without recurrent malignancy who did not receive second-line treatment for chronic GVHD. As expected, patients diagnosed with recurrent malignancy during the first 12 months of initial treatment had poor subsequent survival (Figure 2A). Among patients without recurrent malignancy at 12 months, prior systemic treatment change was not associated with a statistically significant increased risk of subsequent mortality (Figure 2A) but was associated with a statistically significant lower probability of subsequent cure as indicated by complete resolution of chronic GVHD and withdrawal of all systemic treatment (Figure 2B and Table 3).

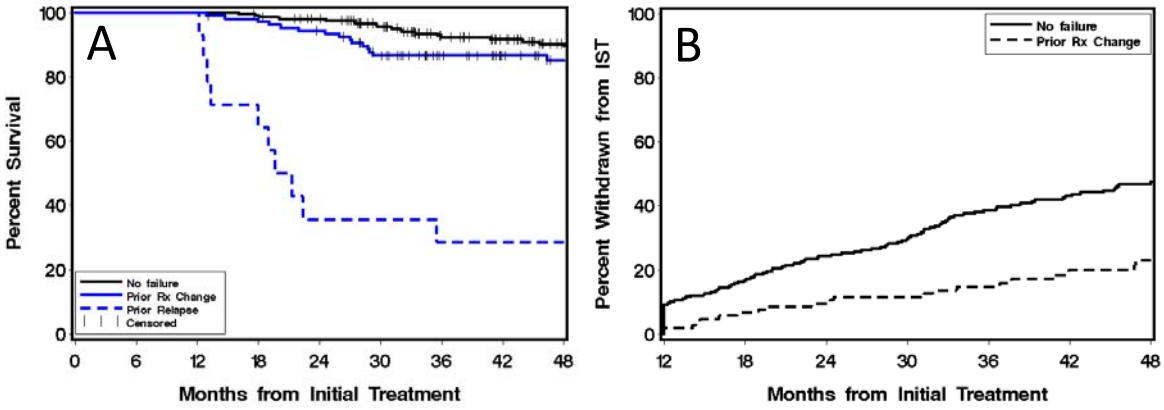


Figure 2. Landmark analysis of outcomes among patients surviving at 12 months after initial systemic treatment for chronic GVHD. Panel A shows survival for patients without recurrent malignancy and no prior change of systemic treatment (—) and for patients without recurrent malignancy but with a prior change of systemic treatment (—). The figure also shows survival for patients with recurrent malignancy diagnosed during the first 12 months of initial systemic treatment for chronic GVHD (– – –). Tic marks show end of follow-up. Panel B shows the cumulative incidence of permanent withdrawal of all systemic treatment after resolution of chronic GVHD according to the presence or absence of a prior systemic treatment change during initial treatment. Patients with recurrent malignancy diagnosed during the first 12 months of initial systemic treatment for chronic GVHD were excluded from this analysis, and death before resolution of chronic GVHD and withdrawal of all systemic treatment was treated as a competing risk.

Table 3. Landmark analysis of outcomes after initial treatment of chronic GVHD, according to FFS*

Endpoints and comparisons	Hazard ratio	95% CI	P
Mortality after 12-month landmark			
FFS (no treatment change before landmark)	1.0 (reference)		
Treatment change before landmark	1.26	0.7 – 2.4	0.48
Discontinued immunosuppression after 12-month landmark			
FFS (no treatment change before landmark)	1.0 (reference)		
Treatment change before landmark	0.41	0.3 – 0.6	<0.001

*patients with relapse before the landmark are excluded in both analyses

Changes of chronic GVHD activity and symptom burden in patients with failure-free survival

As a further test of whether failure-free survival might indicate clinical benefit, changes in measures of GVHD activity and symptom burden were evaluated in patients who participated in a prospective, longitudinal observational study of chronic GVHD (Appendix VII).[18] Cases were included if patients enrolled in the study within 3 months after the initial diagnosis of chronic GVHD, if they had failure-free survival at 12 months, and if data were available at both baseline and at 12 months.

The provider and patient instruments used to collect data are provided in Appendices VIII and IX for reference. The patient instrument includes the Medical Outcomes Study Short Form 36-item Questionnaire version 2 (SF-36v2), Functional Assessment of Chronic Illness Therapy (FACIT), Human

Activity Profile (HAP) and the Lee Symptom Scale. The SF-36v2 has had wide application and is well accepted as a measure of self-reported general health and the degree to which health impairments interfere with activities of daily living and role function.[19,20] The FACIT is an oncology-specific quality-of-life instrument that has well-developed psychometric properties, and population norms for those with both mild and severe chronic illnesses. An additional 18-item disease-specific module evaluates concerns common to patients who have had hematopoietic cell transplantation (FACT-BMT).[21] The HAP presents 94 questions in ascending order according to the metabolic equivalents of oxygen consumption required to perform each activity.[22] The HAP therefore provides a survey of activities that the patient performs independently across a wide range of metabolic demand, beginning with getting out of bed, bathing, dressing, performing a series of progressively more physically demanding household chores, and ending with running or jogging 3 miles in 30 minutes or less. The Lee Symptom Scale is a 30-item, 7-domain symptom scale that has proven reliable, valid and sensitive to change in patients with chronic GVHD (Appendix X).[23]

Sixty-seven patients met criteria for failure-free survival at 12 months. Supplementary Tables 2 and 3 summarizing statistical changes in all measures are provided at the end of this briefing document. Global ratings in Supplementary Table 2 include assessments of overall severity by the provider and patient, measures of general or organ-specific symptom burden, FACT and SF-36 quality of life measures, and performance measures from the Human Activity Profile and Karnofsky score. Provider assessments were available for all 67 patients, and patient-reported items were available from 45 – 48 participants. No item showed statistically significant worsening, and most items in Supplementary Table 2 showed statistically significant improvement. Exceptions included the Lee Symptom Scale components for the lungs, eyes and mouth, and the SF-36 scales for bodily pain, general health perceptions, vitality and mental health.

Measures in Supplementary Table 3 are focused on specific GVHD manifestations. In this analysis, patients were excluded when a manifestation was absent at both baseline and at 1 year. No item showed statistically significant worsening, although a trend suggested worsened range of motion in the ankles ($p = 0.11$). Thirteen of the 31 items showed statistically significant improvement. These included manifestations in the skin, mouth and upper gastrointestinal tract.

The statistical analyses in Supplementary Tables 2 and 3 show general evidence of improvement, but the data do not indicate whether the magnitude of improvement can be interpreted as clinical benefit. The Response Criteria Working Group of the NIH Consensus Project has recommended a set of chronic GVHD-specific core measures for assessing responses in chronic GVHD clinical trials.[24] (See Table 2 in Appendix XI.) Measures include selected provider assessments, symptom scales for providers and patients, and global rating scales for providers and patients. The working group report has also recommended thresholds of change that qualify as complete response, partial response and progression, as summarized in Table 4 below.

To assess the extent to which failure-free survival at 1 year might represent clinical benefit, these criteria defining improvement and worsening of chronic GVHD manifestations were used for a

categorical analysis of changes in the 67 patients with failure-free survival in the prospective, longitudinal observational study of chronic GVHD (Table 5).

Table 4. Core measures of response in chronic GVHD clinical trials

Measure	Improved	Worse
Provider assessments		
NIH Skin Score (0-3)	↓≥ 1	↑≥ 1*
NIH Eye Score (0-3)	↓≥ 1	↑≥ 1*
Modified Oral Mucosa Rating Scale (0-12)	↓≥ 2	↑≥ 2
Total serum bilirubin	↓≥ 50%	↑≥ 2 x ULN
Alanine aminotransferase	↓≥ 50%	↑≥ 2 x ULN
Alkaline phosphatase	↓≥ 50%	↑≥ 2 x ULN
Percent predicted FEV1	↑≥ 10%	↓≥ 10%
NIH Joint and Fascia Score (0-3)	↓≥ 1	↑≥ 1
Photographic range of motion (4-25)	↑≥ 1	↓≥ 1
Provider grading of symptoms		
NIH Lung Symptom Score (0-3)	↓≥ 1	↑≥ 1*
Upper GI Score (0-3)	↓≥ 1	↑≥ 1*
Lower GI Score (0-3)	↓≥ 1	↑≥ 1*
Esophagus Score (0-3)	↓≥ 1	↑≥ 1*
Patient grading of specific symptoms		
Skin itching (0-10)	↓≥ 2	↑≥ 2
Oral sensitivity (0-10)	↓≥ 2	↑≥ 2
Chief eye complaint (0-10)	↓≥ 2	↑≥ 2
Global rating scales		
Provider 0-3	↓≥ 1	↑≥ 1
Provider 0-10	↓≥ 2	↑≥ 2
Patient 0-3	↓≥ 1	↑≥ 1
Patient 0-10	↓≥ 2	↑≥ 2
Lee Symptom Scale (0-100)	↓≥ 7	↑≥ 7

*Changes from 0 to 1 are not counted as progression, since many of these are trivial in magnitude.
ULN, upper limit of normal

Table 5. Categorical changes in patients with failure-free survival at 1 year after enrollment (N = 67)

Measure	Improved	No change	Worse	Unaffected	P*
Provider assessments, N (%)					
NIH Skin Score (0-3)	27 (40)	9 (13)	4 (6)	27 (40)	<0.001
NIH Eye Score (0-3)	13 (19)	18 (27)	8 (12)	28 (42)	0.28
Modified Oral Mucosa Rating Scale (0-12)	26 (39)	23 (34)	8 (12)	10 (15)	0.002
Total serum bilirubin	6 (9)	0 (0)	0 (0)	60 (91)	0.01
Alanine aminotransferase	25 (38)	7 (11)	0 (0)	34 (52)	<0.001
Alkaline phosphatase	18 (27)	15 (23)	1 (2)	32 (48)	<0.001
Percent predicted FEV1	0 (0)	1 (4)	0 (0)	27 (96)	NA
NIH Joint and Fascia Score (0-3)	9 (13)	6 (9)	9 (13)	43 (64)	1.00
Photographic range of motion (4-25)	5 (11)	5 (11)	13 (29)	22 (49)	0.06
Provider grading of symptoms, N (%)					
NIH Lung Symptom Score (0-3)	11 (16)	8 (12)	0 (0)	48 (72)	0.001
Upper GI Score (0-3)	15 (22)	1 (1)	0 (0)	51 (76)	<0.001
Lower GI Score (0-3)	6 (9)	3 (4)	1 (1)	57 (85)	0.06
Esophagus Score (0-3)	6 (9)	2 (3)	1 (1)	58 (87)	0.06
Patient grading of specific symptoms, N (%)					
Skin itching (0-10)	16 (36)	14 (31)	5 (11)	10 (22)	0.02
Oral sensitivity (0-10)	19 (40)	9 (19)	4 (9)	15 (32)	0.002
Chief eye complaint (0-10)	14 (30)	11 (24)	13 (28)	8 (17)	0.85
Global rating scales, N (%)					
Provider 0-3	36 (54)	27 (40)	4 (6)	NA	<0.001
Provider 0-10	38 (57)	27 (40)	2 (3)	NA	<0.001
Patient 0-3	17 (36)	25 (53)	4 (9)	1 (2)	0.005
Patient 0-10	21 (50)	14 (33)	4 (10)	3 (7)	0.001
Lee Symptom Scale (0-100)	13 (28)	27 (57)	7 (15)	NA	0.18

*binomial test for equal proportions between improved and worsened patients

Seven of 13 provider assessments showed statistically significant improvement in the skin, mouth, lung, upper gastrointestinal tract and liver. Both global ratings by providers also showed statistically significant improvement. Patient reported assessments showed statistically significant improvement in the skin and mouth but not the eyes. Both global ratings by patients showed statistically significant improvement, but the Lee Symptom Scale did not show statistically significant evidence of improvement in the categorical analysis. The photographic range of motion showed a trend suggesting worsening (P = 0.06).

b. To what extent could failure-free survival at 6 months be considered as an indicator of clinical benefit in early phase trials of second-line systemic treatment for chronic GVHD?

This question is very similar to the question above. A large landmark analysis analyzed outcomes among patients who were alive at 6 months after the onset of second-line systemic treatment for chronic GVHD.[25] (See Appendix XII for full details.) Patients diagnosed with recurrent malignancy during the first 6 months of second-line treatment had poor subsequent survival (Figure 3A). Among patients without recurrent malignancy at 6 months, prior systemic treatment change was not associated with a statistically significant increased risk of mortality (Figure 3A) but was associated with a statistically significant lower probability of subsequent cure as indicated by complete resolution of chronic GVHD and withdrawal of all systemic treatment (Figure 3B and Table 6).

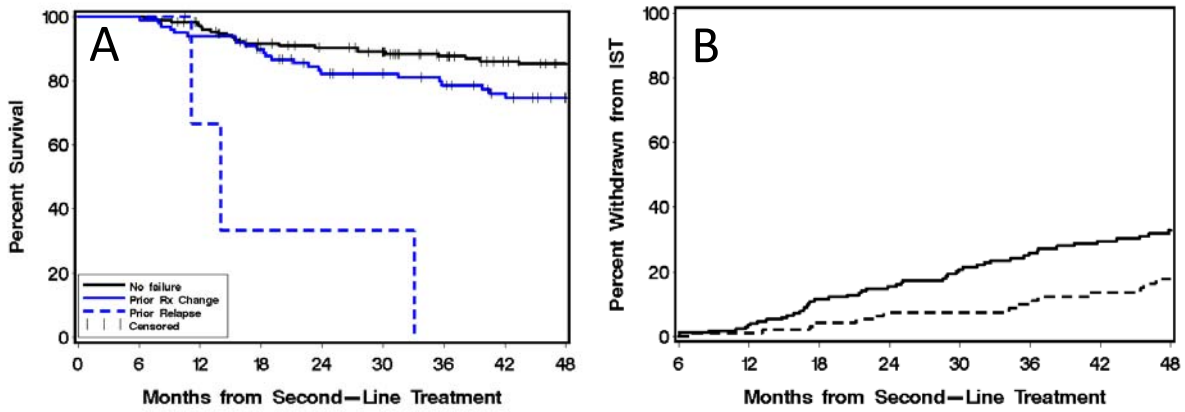


Figure 3. Landmark analysis of outcomes among patients surviving at 6 months after second-line systemic treatment for chronic GVHD. Panel A shows survival for patients without recurrent malignancy and no prior change of systemic treatment (—) and for patients without recurrent malignancy but with a prior change of systemic treatment (—). The figure also shows survival for patients with recurrent malignancy diagnosed during the first 6 months of second-line systemic treatment for chronic GVHD (- - -). Tic marks show end of follow-up. Panel B shows the cumulative incidence of permanent withdrawal of all systemic treatment after resolution of chronic GVHD according to the presence or absence of a prior systemic treatment change during second-line treatment. Patients with recurrent malignancy diagnosed during the first 6 months of second-line systemic treatment for chronic GVHD were excluded from this analysis, and death before resolution of chronic GVHD and withdrawal of all systemic treatment was treated as a competing risk.

Table 6. Landmark analysis of outcomes after second-line treatment of chronic GVHD, according to FFS*

Endpoints and comparisons	Hazard ratio	95% CI	P
Mortality after 6 month landmark			
FFS (no treatment change before landmark)	1.0 (reference)		
Treatment change before landmark	1.61	1.0 – 2.7	0.07
Discontinued immunosuppression after 6 month landmark			
FFS (no treatment change before landmark)	1.0 (reference)		
Treatment change before landmark	0.53	0.3-0.8	0.005

*patients with relapse before the landmark are excluded in both analyses

c. Should the absence of recurrent or progressive malignancy be included as a component in the definition of failure-free survival?

If distinct treatments might be associated with different risks of recurrent or progressive malignancy in patients who already have chronic GVHD, then these events should be considered as failure in the definition of failure-free survival. On the other hand, if distinct treatments are unlikely to be associated with different risks of recurrent or progressive malignancy in patients who already have chronic GVHD, then these events should be treated as competing risks, such that the outcome of treatment for chronic GVHD cannot be reliably determined.

In the absence of definitive data, we believe that recurrent or progressive malignancy should be included in the definition of failure, since it is possible that excessive immunosuppression could dampen graft-versus-leukemia effects. In the published reports,[17,25] approximately 95% of patients had malignant diseases as the indication for hematopoietic cell transplantation. Nonetheless, only 7% of patients had failure due to recurrent malignancy during the first 12 months after initial treatment. Likewise, only 4% of patients had failure due to recurrent malignancy during the first 6 months after second-line treatment. Hence, the relative contribution of recurrent or progressive malignancy to the overall risk of failure is small. Atypical cohorts comprised predominantly of patients with nonmalignant diseases could have slightly higher rates of failure-free survival than observed in the published reports.[17,25]

d. Should steroid doses below a predefined threshold at 12 months after initial treatment or at 6 months after second-line treatment be included as an additional criterion of failure-free survival?

In theory, it might be possible to avoid changing treatment in trials with failure-free survival as an endpoint by maintaining steroid doses at high levels for extended periods. In the published studies,[17,25] lower steroid doses across a range of thresholds in patients with FFS at 12 months after initial treatment and at 6 months after second-line treatment were consistently associated with higher probabilities of attaining the intended treatment effect of durably controlling the disease without adding any new systemic treatment until withdrawal of all systemic treatment. (See Figure 3 in Appendix VI and Figure 3 in Appendix XII.)

The question here is whether prednisone doses above some threshold at the landmark time point should be considered equivalent to a prior change of systemic treatment, because outcomes for patients with prednisone doses above the threshold are similar to those for patients who had a prior change of systemic treatment. Data addressing this question are shown below in Figure 4 and Table 7 with respect to initial treatment and in Figure 5 and Table 8 with respect to second-line treatment. Threshold prednisone doses were selected to approximate the 50^{th} percentile at the time of the landmark, and the 25^{th} and 75^{th} percentile doses were included as a sensitivity analysis. The term “cure” indicates resolution of chronic GVHD and permanent withdrawal of all systemic treatment.

Initial treatment

The composite endpoint of prior treatment change or FFS with prednisone dose ≥0.05 mg/kg/day at 12 months after initial treatment (50th percentile) was associated with a statistically significant increased risk of subsequent mortality (p = 0.05) (Figure 4A and Table 7) and a statistically significant decreased probability of cure (p = 0.003) (Figure 4B and Table 7). The composite endpoint of prior treatment change or FFS with prednisone at any dose (25th percentile) at 12 months after initial treatment was not associated with a statistically significant increased risk of subsequent mortality (p = 0.16) or a statistically significant decreased probability of cure (p = 0.08) (Figure 4C and 4D, and Table 7). The composite endpoint of prior treatment change or FFS with prednisone dose ≥0.2 mg/kg/day (75th percentile) was not associated with a statistically significant increased risk of subsequent mortality (p = 0.25) or a statistically significant decreased probability of cure (p = 0.07) (Figure 4E and 4F, and Table 7).

Second-line treatment

The composite endpoint of prior treatment change or FFS with prednisone dose ≥0.2 mg/kg/day (50th percentile) at 6 months after second-line treatment was associated with a statistically significant increased risk of subsequent mortality (p = 0.03) (Figure 5A and Table 8) and a statistically significant decreased probability of cure (p = 0.02) (Figure 5B and Table 8). Results were similar with the threshold prednisone dose set at 0.1 mg/kg/day (25th percentile) (Figure 5D and 5E and Table 8). The composite endpoint of prior treatment change or FFS with prednisone dose ≥0.3 mg/kg/day (75th percentile) was associated with a statistically significant increased risk of subsequent mortality (p = 0.006) (Figure 5E and Table 8) but not with a statistically significant decreased probability of cure (p = 0.12) (Figure 5F and Table 8).

Discussion of failure-free survival as an endpoint for chronic GVHD treatment trials

The strengths and weaknesses of FFS have been addressed extensively in the Discussion of published reports.[17,25] Both reports were accompanied by favorable editorials commenting on potential utility of using FFS as an endpoint in the interim until methods for measuring global response have been developed and validated for use as the primary endpoint in clinical trials.[26,27] (See Appendices XIII and XIV.) The absence of non-relapse mortality and recurrent malignancy as components of FFS each reflect clinical benefit. In the published studies, the absence of systemic treatment change at 12 months after initial treatment and at 6 months after second-line treatment was associated with a higher probability of cure of chronic GVHD. The absence of prior treatment change was not associated with a statistically significant improvement in survival after 12 months of initial treatment or after 6 months of second-line treatment.

The data summarized above suggest that the absence of systemic treatment change combined with a threshold prednisone dose as a criterion of failure could reflect clinical benefit. First, as mentioned above, lower steroid doses across a range of thresholds in patients with FFS at 12 months after initial treatment and at 6 months after second-line treatment were consistently associated with higher probabilities of attaining the intended effect of durably controlling the disease (i.e., without adding any

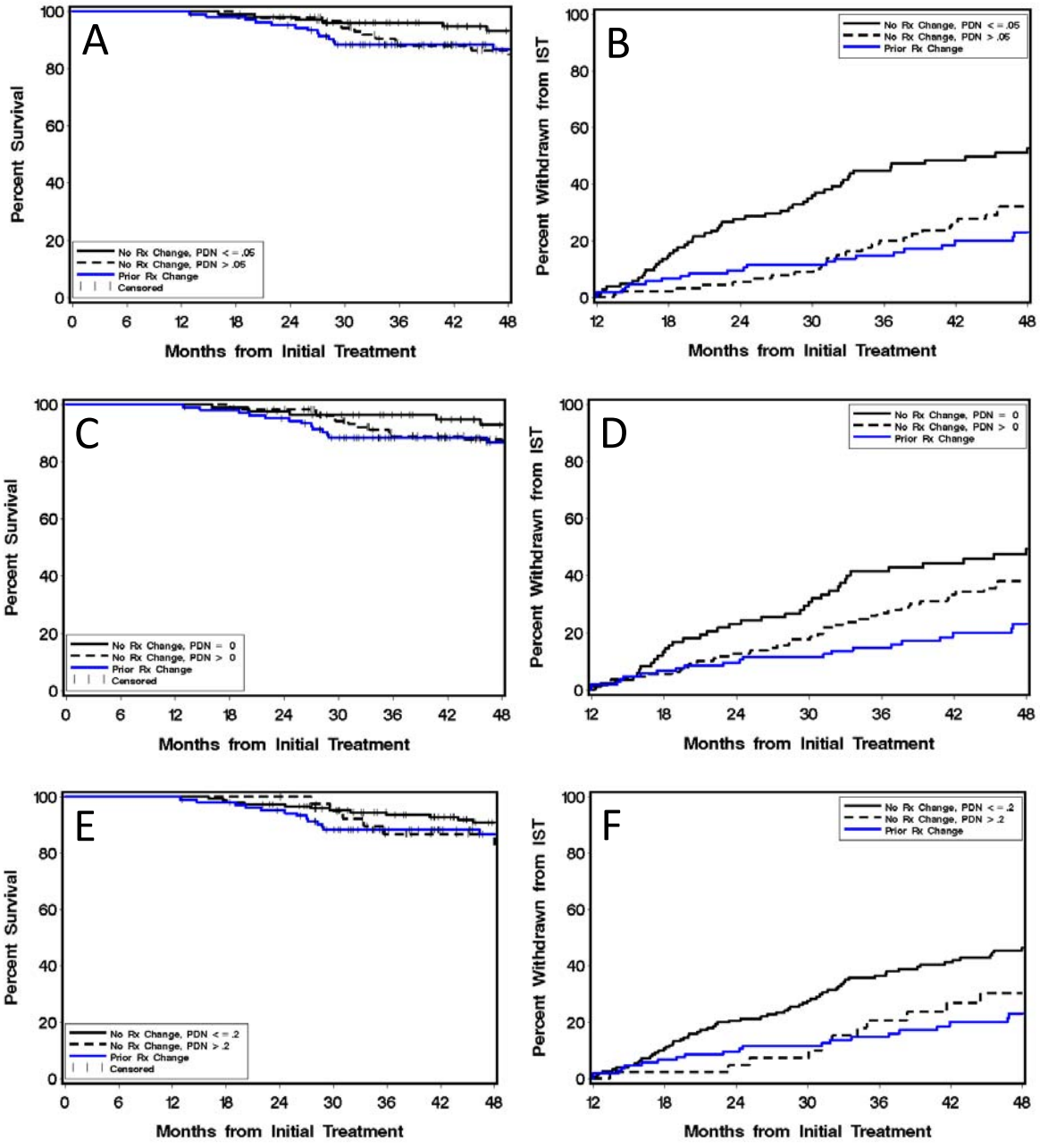


Figure 4. Landmark analysis showing the effect of adding prednisone dose at 12 months as an additional criterion for failure-free survival after initial systemic treatment of chronic GVHD. Panels A, C and E show survival after the 12-month landmark. Groups are defined according to the presence of prior systemic treatment change (—), or the absence of prior treatment change with prednisone doses at 6 months ≤0.05 (**—**) vs. >0.05 (**- - -**) mg/kg/day (~50th percentile) (A), no prednisone (**—**) vs. any prednisone (**- - -**) mg/kg/day (~25th percentile) (C), or ≤0.2 (**—**) vs. >0.2 (**- - -**) mg/kg/day (75th percentile) (E). Tic marks show end of follow-up. Panels B, D and F show the cumulative incidence of withdrawal from immunosuppressive treatment (IST) for the same groups. Relapse and death are competing risks for withdrawal from immunosuppressive treatment. Patients diagnosed with recurrent malignancy during the first 12 months of initial treatment were excluded from all analyses.

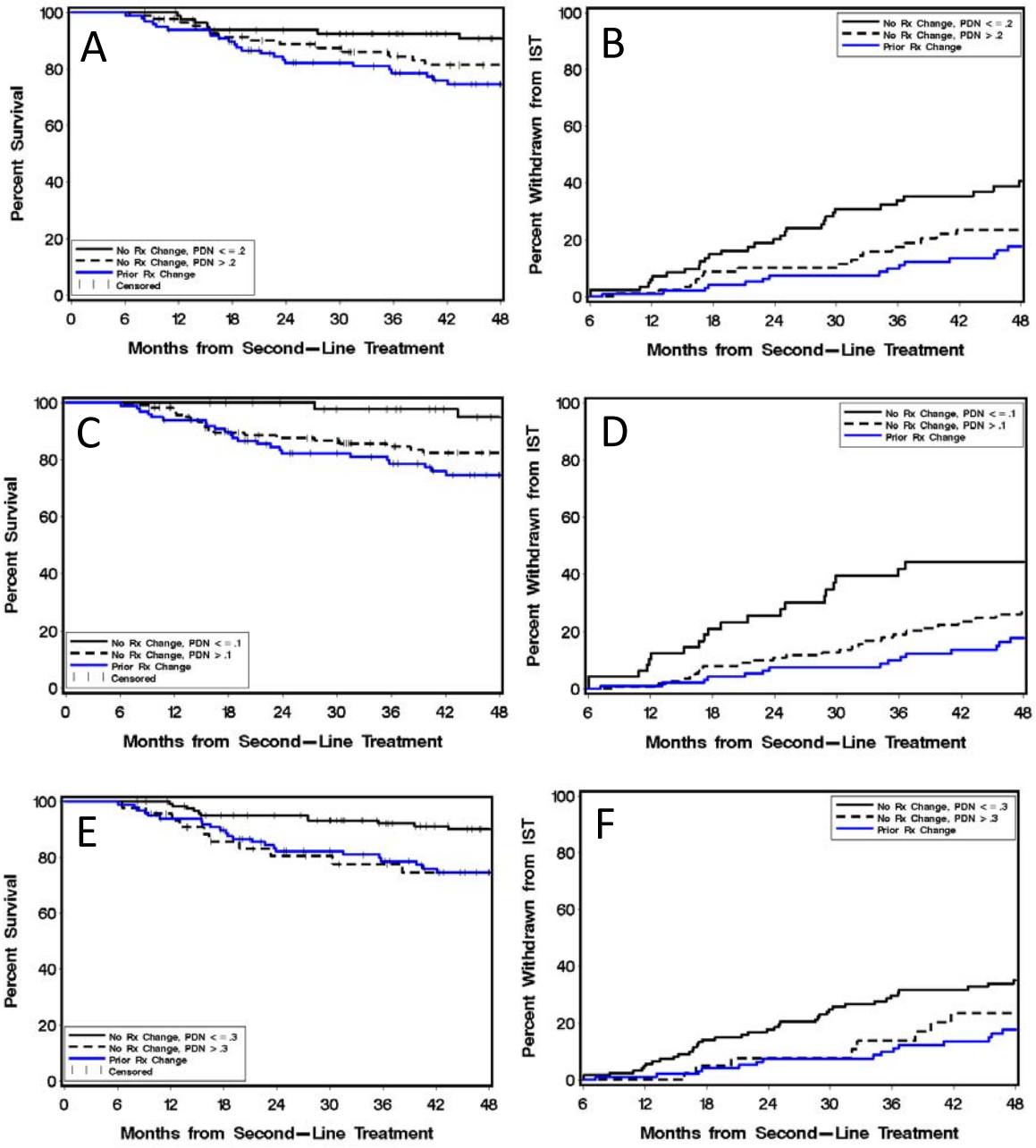


Figure 5. Landmark analysis showing the effect of adding prednisone dose at 6 months as an additional criterion for failure-free survival after second-line systemic treatment of chronic GVHD. Panels A, C and E show survival after the 6-month landmark. Groups are defined according to the presence of prior systemic treatment change (—), or the absence of prior treatment change with prednisone doses at 6 months ≤0.2 (**—**) vs. >0.2 (**- - -**) mg/kg/day (~50th percentile) (A), ≤0.1 (**—**) vs. >0.1 (**- - -**) mg/kg/day (~25th percentile) (C), or ≤0.3 (**—**) vs. >0.3 (**- - -**) mg/kg/day (~75th pecentile) (E). Tic marks show end of follow-up. Panels B, D and F show the cumulative incidence of withdrawal from immunosuppressive treatment (IST) for the same groups. Relapse and death are competing risks for withdrawal from immunosuppressive treatment. Patients diagnosed with recurrent malignancy during the first 6 months of second-line treatment were excluded from all analyses.

Table 7. Landmark analysis of outcomes after initial treatment of chronic GVHD, according to FFS with a threshold prednisone dose added to define failure*

Endpoints and comparisons	Hazard ratio	95% CI	P
Mortality after 12-month landmark			
FFS with prednisone dose ≤0.05 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.05 mg/kg/day	2.31	1.0 – 5.4	0.05
FFS, no treatment with prednisone	1.0 (reference)		
Prior treatment change or FFS with any prednisone treatment	1.87	0.8 – 4.5	0.16
FFS with prednisone dose ≤0.2 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.2 mg/kg/day	1.64	0.7 – 3.8	0.25
Discontinued immunosuppression after 12-month landmark			
FFS with prednisone dose ≤0.05 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.05 mg/kg/day	0.53	0.3 – 0.8	0.003
FFS, no treatment with prednisone	1.0 (reference)		
Prior treatment change or FFS with any prednisone treatment	0.69	0.5 – 1.0	0.08
FFS with prednisone dose ≤0.2 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.2 mg/kg/day	0.58	0.3 – 1.0	0.07

*patients with relapse before the landmark are excluded in all analyses

Table 8. Landmark analysis of outcomes after second-line treatment of chronic GVHD, according to FFS with a threshold prednisone dose added to define failure*

Endpoints and comparisons	Hazard ratio	95% CI	P
Mortality after 6-month landmark			
FFS with prednisone dose ≤0.3 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.3 mg/kg/day	2.69	1.3 – 5.5	0.006
FFS with prednisone dose ≤0.1 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.1 mg/kg/day	6.22	1.5 – 26	0.01
FFS with prednisone dose ≤0.2 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.2 mg/kg/day	2.41	1.1 – 5.2	0.03
Discontinued immunosuppression after 6 month landmark			
FFS with prednisone dose ≤0.2 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.2 mg/kg/day	0.57	0.4 – 0.9	0.02
FFS with prednisone dose ≤0.1 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.1 mg/kg/day	0.53	0.3 – 0.9	0.01
FFS with prednisone dose ≤0.3 mg/kg/day	1.0 (reference)		
Prior treatment change or FFS with prednisone dose >0.3 mg/kg/day	0.60	0.3 – 1.1	0.12

*patients with relapse before the landmark are excluded in all analyses

new systemic treatment) until withdrawal of all systemic treatment. Second, prednisone doses lower than the 50th percentile among patients with FFS at 12 months after initial treatment and at 6 months after second-line treatment were associated with better subsequent survival. This association was consistent across a range of threshold prednisone doses after second-line treatment but not after initial treatment. Third, prednisone doses lower than the 50th percentile among patients with FFS at 12 months after initial treatment and at 6 months after second-line treatment were associated with a higher probability of subsequent cure of chronic GVHD. This association did not reach statistical significance

across all threshold prednisone doses after either initial treatment or second-line treatment. The different thresholds used in the sensitivity analysis shift relatively small numbers of patients from one group to the other, and differences in hazard ratio estimates, 95% confidence intervals and p-values reflect the random variation embodied in small numbers as well as systematic effects. Therefore, the data do not identify the optimal threshold prednisone dose for the endpoint criterion.

Concerns have been raised about the reliability of using treatment change as an element in the composite FFS endpoint. Failure-free survival can be measured objectively in the sense that each of the composite elements can be verified, but the dosing of prednisone and decisions to change systemic treatment depend on subjective clinical judgment. Failure-free survival measures benefit as the absence of new harm caused by the disease, and the absence of harm is inferred indirectly from the absence of a new systemic treatment change. Introduction of a new systemic treatment in a patient with chronic GVHD can be taken as evidence that current treatment has not provided satisfactory results, because chronic GVHD manifestations are progressing, persisting without improvement, or improving more slowly than desired. This interpretation can be confounded, however, when changes are motivated by toxicity, inconvenience, financial burden, or the availability of a newer alternative that is more attractive to the patient or physician.

These same concerns also apply in the assessment of response, since consensus has been established that if additional systemic therapy for chronic GVHD is added before the end of the specified study period, the outcome is categorized as lack of response. In a randomized trial with a crossover option, for example, some patients who were evaluated as having a "response" crossed over to the other arm. This pairing between assessment and action might appear to be inconsistent, but the baseline for comparison in clinical practice is typically the most recent clinic visit, not the entry point into the study. A systemic treatment change might be perfectly reasonable for a patient with chronic GVHD that has improved when compared to enrollment but worsened when compared to the most recent visit. For assessment of failure-free survival and response in future clinical trials, these problems could be mitigated by implementing clear guidelines for changing therapy, by requiring clear documentation of reasons for beginning new systemic treatment and by blinding in order to minimize the risk of bias.

The published reports[17,25] represent data from retrospective studies at a single center. Studies have not yet been carried out to determine whether the reported data are representative of results from other centers, although results from a prospective, multicenter study of initial treatment for chronic GVHD were similar to those in the published report.[17] Paired comparisons from the prospective longitudinal observational study have suggested that patients with FFS at 1 year have measurable overall reductions in symptom burden, disease activity and functional impairment, although the extent to which response shift might account for these results is not known. Results from the BMT CTN 0801 trial (NCT01106833) could determine whether these observations hold true in the context of a prospective clinical trial for chronic GVHD. This trial was designed to evaluate a regimen of steroids and sirolimus with or without a calcineurin inhibitor in patients at onset of systemic treatment or in patients who had an inadequate response within 3 months after initial systemic treatment.

The primary endpoint has paramount importance in evaluating the results of clinical trials, but success cannot be evaluated based on the primary endpoint alone. Trials using FFS as the primary endpoint must include a variety of secondary endpoints that individually assess changes in symptom burden, disease activity and disease-related impairment, even if no global response measure of response has been fully developed and validated. Measures made at baseline and at the time of endpoint assessment should be used to derive change scores for each patient. In randomized trials, change scores should be compared between arms in order to determine whether improvements in FFS are supported by evidence of reduced symptom burden, decreased disease activity, and an absence of emerging damage or impairment. In single-arm studies, paired statistical tests should be used to determine whether patients with FFS have measurable reductions in symptom burden, disease activity and functional impairment.

Survival without progressive impairment

e. To what extent could prevention of "progressive impairment" at 2 years be considered as an indicator of clinical benefit in late phase trials of treatment for chronic GVHD?

As discussed above, treatment of chronic GVHD is intended to produce a sustained benefit by reducing symptom burden, controlling objective manifestations of disease, and preventing organ damage and progressive impairment leading to disability, while avoiding disproportionate toxicity related to treatment. The term "progressive impairment" is intended to capture the emergence of an enduring chronic GVHD-related health state that threatens or compromises a patient's physical well-being or function in ways that cannot be easily reversed. Hence, "progressive impairment" indicates inadequately controlled chronic GVHD. In the following sections, we propose a provisional definition of impairment, evaluate its application in a prospective cohort study, and discuss the merits of survival without progressive impairment as an indicator of clinical benefit in clinical trials.

Questions for discussion:

- Should any items be omitted from or added to the list of outcomes categorized as progressive impairment, or should any of the proposed item thresholds be modified?
- What additional analyses would help determine whether this endpoint represents clinical benefit?
- Given the typical time course of chronic GVHD, what would be the most appropriate time point for comparison between arms in a late-phase controlled trial using survival without progressive impairment as the primary endpoint?
- Given the typical time course of chronic GVHD, could earlier time points be used for comparison between arms in earlier-phase controlled trials using survival without progressive impairment as the primary endpoint?

Definition of impairment

We selected 21 items as having face validity indicating progressive impairment. This list was drawn from items routinely collected in a longitudinal, prospective multicenter study of patients with chronic GVHD.[18] These items fall in 3 broad categories: providers' assessments, patient surveys and the forced

expiratory volume in the first second (FEV1) component of pulmonary function tests. The items selected are summarized below and annotated with pertinent sections of the instruments used to collect the data. Complete copies of the instruments used for provider assessments and patient surveys are provided in Appendices VIII and IX.

i. Measures from Chronic GVHD Provider Assessments

1) Absolute increase in the sum of moveable and non-moveable sclerosis by >20%. A 20% absolute difference approximates the minimum change that can be reliably detected by physical examination.

Do not use Rule of 9s *Indicate % of body part affected*	Erythematous rash of any sort	Moveable sclerosis	Non-moveable subcutaneous sclerosis or fasciitis
1. Head/neck/scalp	%	%	%
2. Anterior torso	%	%	%
3. Posterior torso	%	%	%
4. L. upper extremity	%	%	%
5. R. upper extremity	%	%	%
6. L. lower extremity, (incl. L buttock)	%	%	%
7. R. lower extremity, (incl. R buttock)	%	%	%
8. Genitalia ❑ not examined	%	%	%

2) Any increase in the global measure of skin sclerosis

	0	1	2	3	4
Skin sclerotic changes	❑ Normal	❑ Thickened with pockets of normal skin	❑ Thickened over majority of skin	❑ Thickened, unable to move	❑ Hidebound, unable to pinch

3) Increase in the global skin score from 0 – 1 to 2 or 3, or from 2 to 3

	0	1	2	3
Skin Score	❑ No Symptoms	❑ <18% BSA with disease signs but NO sclerotic features	❑ 19-50% BSA OR involvement with superficial sclerotic features "not hidebound" (able to pinch)	❑ >50% BSA OR deep sclerotic features "hidebound" (unable to pinch) OR impaired mobility, ulceration or severe pruritus
Fascia	❑ Normal	❑ Tight with normal areas	❑ Tight	❑ Tight, unable to move

4) Any increase in fascia score (see above)

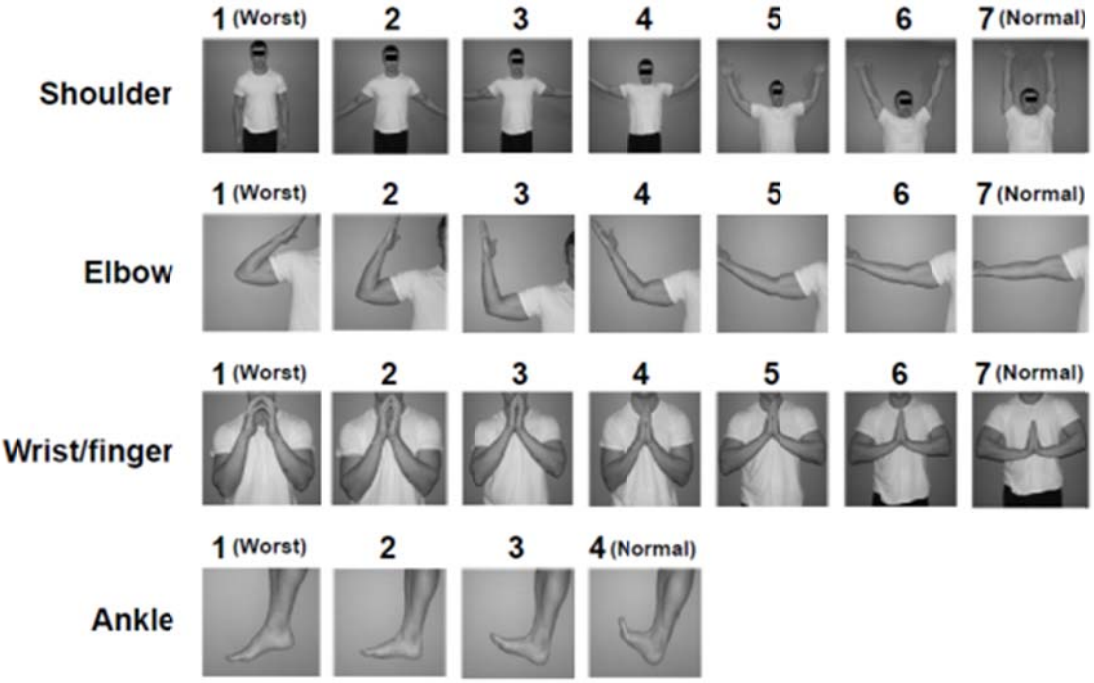
5) Absolute increase in the sum of grade 3 – 4 skin involvement by >20%. A 20% absolute difference approximates the minimum change that can be reliably detected by physical examination.

SKIN

Region	Grade	% Area of Grade	Fraction of Grade 3 or 4 Areas with Erythema (indicate up to what fraction is involved)	Region	Grade	% Area of Grade	Fraction of Grade 3 or 4 Areas with Erythema (indicate up to what fraction is involved)
1. Head, Neck and Scalp	0	%		6. Right Hand	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	100 %			Total =	100 %	
2. Chest	0	%		7. Left Arm	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	100 %			Total =	100 %	
3. Abdomen and Genitals	0	%		8. Left Hand	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	100 %			Total =	100 %	
4. Back and Buttocks	0	%		9. Right Leg and Foot	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	100 %			Total =	100 %	
5. Right Arm	0			10. Left Leg and Foot	0		
	1				1		
	2				2		
	3		❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	100 %			Total =	100 %	

Percentages must add up to 100
0 = normal skin
1 = discolored [hypopigmentation, hyperpigmentation, alopecia, erythema, maculopapular rash]
2 = lichenoid plaque, or skin thickened (able to move)
3 = skin thickened with limited motion but able to pinch [scleroderma or fasciae involvement]
4 = hidebound skin, unable to move, unable to pinch

6) Decrease of any joint range of motion score by ≥2 points, indicating increasingly restricted range of motion in shoulder and elbow extension, wrist and ankle dorsiflexion, and finger extension



7) Increase of oral ulceration, esophagus, upper GI, lower GI, eye, joint/fascia, genital, or lung score from 0 – 2 to 3 (see following pages)

		0	1	2	3
Mouth	Ulcers	❑ None	❑ None	❑ Ulcers involving (≤20%)	❑ Severe ulcerations (>20%)

		0	1	2	3
Gastro-intestinal	Esophagus • Dysphagia OR • Odynophagia	❑ No esophageal symptoms	❑ Occasional dysphagia or odynophagia with solid food or pills *during the past week*	❑ Intermittent dysphagia or odynophagia with solid food or pills (but not for liquids or soft foods) *during the past week*	❑ Dysphagia or odynophagia for almost all oral intake, *on almost every day of the past week*
	Upper GI • Early satiety OR • Anorexia OR • Nausea & vomiting	❑ No symptoms	❑ Mild, occasional symptoms with little reduction in oral intake *during the past week*	❑ Moderate, intermittent symptoms throughout the day, with some reduction in oral intake, *during the past week*	❑ More severe or persistent symptoms throughout the day, with marked reduction in oral intake, *on almost every day of the past week*
	Lower GI • Diarrhea	❑ No loose or liquid stools *during the past week*	❑ Occasional loose or liquid stools, on some days *during the past week*	❑ Intermittent loose or liquid stools throughout the day, *on almost every day of the past week* **without requiring** intervention to prevent or correct volume depletion	❑ Voluminous diarrhea *on almost every day of the past week* **requiring** intervention to prevent or correct volume depletion

	0	1	2	3
Eye Score	❑ No symptoms	❑ Mild dry eye symptoms not affecting ADL (requiring eye drops <3x per day) OR asymptomatic signs of kerato-conjunctivitis sicca	❑ Moderate dry eye symptoms partially affecting ADL (requiring eye drops >3x per day or punctual plugs) WITHOUT vision impairment	❑ Severe dry eye symptoms significantly affecting ADL (special eyewear to relieve pain) OR unable to work because of ocular symptoms OR loss of vision caused by kerato-conjunctivitis sicca
Joints and Fascia Score	❑ No symptoms	❑ Mild tightness of arms or legs, normal or mild decreased range of motion (ROM) AND not affecting ADL	❑ Tightness of arms or legs OR joint contractures, erythema thought due to fasciitis, moderate decrease ROM AND mild to moderate limitation of ADL	❑ Contracture WITH significant decrease of ROM AND significant limitation of ADL (unable to tie shoes, button shirts, dress self etc.)
Genital Tract Score **(score even if no GYN exam, required for men too)** ❑ No GYN Exam	❑ No symptoms	❑ Symptomatic with mild distinct signs on exam AND no effect on coitus and minimal discomfort with GYN exam	❑ Symptomatic with distinct signs on exam AND with mild dyspareunia or discomfort with GYN exam	❑ Symptomatic WITH advanced signs (stricture, labia agglutination or severe ulceration) AND severe pain with coitus or inability to insert vaginal spectrum
Lung Score	❑ No symptoms	❑ Mild symptoms (shortness of breath after climbing one flight of steps)	❑ Moderate symptoms (shortness of breath after walking on flat ground)	❑ Severe symptoms (shortness of breath at rest; requiring O_2)

8) Increase of oral pain (page 4)score or GI global score from 0 – 1 to 2 – 3 or from 2 to 3

	0	1	2	3
Mouth Pain	❑ No symptoms	❑ Food sensitivity	❑ Pain requiring narcotics	❑ Unable to eat

	0	1	2	3
GI Tract Score	❑ No symptoms	❑ Symptoms such as dysphagia, anorexia, nausea, vomiting, abdominal pain or diarrhea without significant weight loss (<5%)	❑ Symptoms associated with mild to moderate weight loss (5-15%)	❑ Symptoms associated with significant weight loss >15%, requires nutritional supplement for most calorie needs OR esophageal dilation

9) Increase of any other chronic GVHD indicator from 0 – 3 to 4

Other indicators, clinical manifestations or severe complications related to chronic GVHD					
	Never (0)	Past, not now (1)	Mild (2)	Moderate (3)	Severe (4)
1. Pleural Effusion(s)	❑	❑	❑	❑	❑
2. Bronchiolitis obliterans	❑	❑	❑	❑	❑
3. Bronchiolitis obliterans organizing pneumonia	❑	❑	❑	❑	❑
4. Nephrotic syndrome	❑	❑	❑	❑	❑
5. Malabsorption	❑	❑	❑	❑	❑
6. Esophageal stricture or web	❑	❑	❑	❑	❑
7. Ascites (serositis)	❑	❑	❑	❑	❑
8. Myasthenia Gravis	❑	❑	❑	❑	❑
9. Peripheral Neuropathy	❑	❑	❑	❑	❑
10. Polymyositis	❑	❑	❑	❑	❑
11. Pericardial Effusion	❑	❑	❑	❑	❑
12. Cardiomyopathy	❑	❑	❑	❑	❑
13. Cardiac conduction defects	❑	❑	❑	❑	❑
14. Coronary artery involvement	❑	❑	❑	❑	❑
15. Other, please specify: _____	❑	❑	❑	❑	❑
16. Other, please specify: _____	❑	❑	❑	❑	❑
17. Other, please specify: _____	❑	❑	❑	❑	❑

ii. Measures from the Chronic GVHD Patient Survey

1) Decrease in KPS from 1 – 6 to 7 – 9

Which statement describes how you feel most of the time? (please check one)

- ☐ 1. Normal, no difficulties with daily activities
- ☐ 2. Able to carry on normal activities, minor problems
- ☐ 3. Normal activity with effort
- ☐ 4. Able to care for self, but unable to carry on normal activity or active work
- ☐ 5. Require occasional assistance, but able to care for most of needs
- ☐ 6. Require considerable assistance and frequent medical care
- ☐ 7. Disabled, require special care and assistance
- ☐ 8. Severely disabled, hospitalized
- ☐ 9. Very sick, hospitalized

2) Decrease in the physical function subscale of the SF-36 by ≥ 0.5 standard deviation. The complete instrument is included in Appendix IX. The standard deviation is defined according to population norms. Changes ≥0.5 standard deviation are generally considered clinically meaningful.

3) Decrease in the Human Activity Profile (HAP) by ≥0.5 standard deviation. The complete instrument is included in Appendix IX. The standard deviation is defined according to population norms. Changes ≥0.5 standard deviation are generally considered clinically meaningful.

4) Decrease Physical Component Summary (PCS) score of the SF-36 by ≥0.5 standard deviation. The complete instrument is included in Appendix IX. The standard deviation is defined according to population norms. Changes ≥0.5 standard deviation are generally considered clinically meaningful.

iii. Measures from chart review

1) Absolute decrease in FEV1 by ≥10% from first measurement, based on chart review.

Patients enrolled as incident cases of chronic GVHD were assessed at 3 and 6 months after enrollment, and then at 6-month intervals. Patients enrolled as prevalent cases were assessed at 6-month intervals after enrollment. In all cases, results were compared to the baseline at enrollment. Except for FEV1, the occurrence of any change listed above in 2 successive evaluations was counted as "progressive impairment." A decrease in FEV1 by ≥10% from baseline at any evaluation was counted as progressive impairment in the analysis summarized below.

Evaluation of survival without progressive impairment

We tested this definition of progressive impairment in a multicenter cohort of 575 patients who had a baseline assessment completed on or before January 31, 2013. Across 1,855 follow-up visits, 237 patients (41%) met criteria for progressive impairment during a median of 39.9 (range, 3.8 to 69.2) months of follow-up, 101 (18%) died without meeting criteria for progressive impairment, and 237 (41%) are surviving without progressive impairment. Of the 101 patients who died without meeting criteria for progressive impairment, 29 had no reported new impairment at any follow-up visit before death. Of the 237 patients surviving without progressive impairment, 102 had at least 1 assessment with impairment reported, but not 2 successive assessments according to the proposed definition of progressive impairment, 109 have had no reported new impairment at any follow-up visit, and 26 have had no follow-up visits.

The cumulative incidence of progressive impairment was 20% at 6 months, 33% at 12 months, and 44% at 24 months (Figure 6A). Survival without progressive impairment was 74% at 6 months, 56% at 12 months, and 38% at 24 months (Figure 6B).

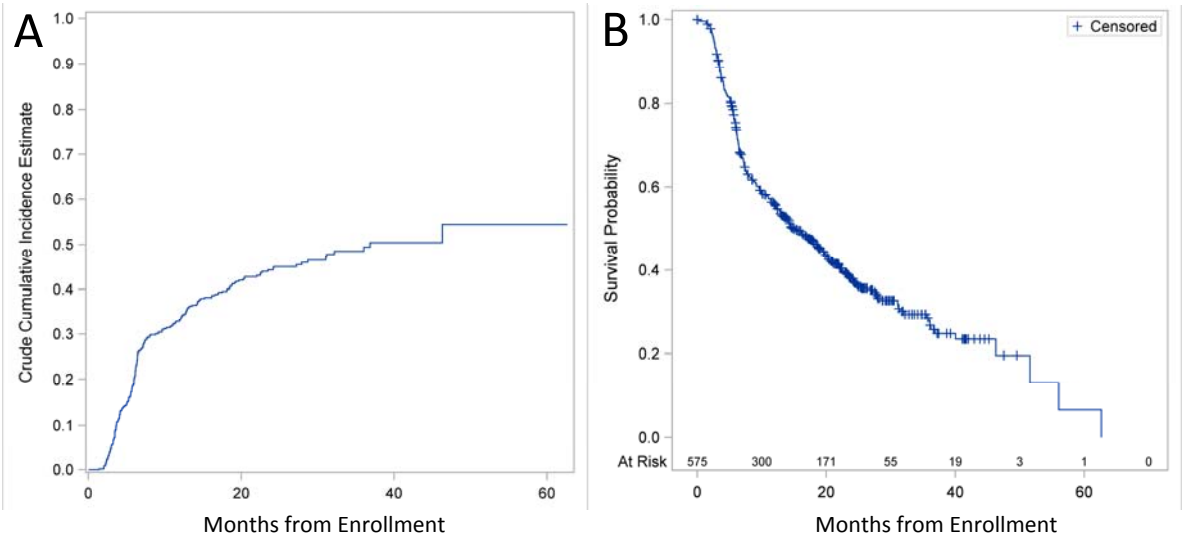


Figure 6. Progressive impairment in a longitudinal prospective study of 575 patients with chronic GVHD. A) Cumulative incidence of progressive impairment with death treated as a competing risk. B) Survival without progressive impairment or death. Tic marks indicate the end of follow-up. In 29 patients, follow-up was truncated by recurrent malignancy.

Of the 237 patients with progressive impairment, 125 (53%) had progressive impairment based on provider measures, 61 (26%) based on patient-reported measures alone, and 44 (19%) based on decreased FEV1 alone (Table 9). In 68 (29%) patients, progressive impairment was based on more than one category.

Table 9. Categories of progressive impairment (N = 237)

Category	Percent
Provider measures alone	27
Patient measures alone	26
FEV1 alone	19
Provider measures + patient measures	13
Provider measures + FEV1	7
Patient measures + FEV1	4
Provider measures + patients measures + FEV1	5

In descending rank order, the most frequent causes contributing to progressive impairment by this definition were decreased FEV1, decreased HAP score, increased global measure of skin sclerosis, increased fascia score, increased global skin score, decreased performance in the SF-36 physical function subscale, and decreased performance in the physical component summary score of the SF-36 (Table 10).

Table 10. Progressive impairment in a longitudinal prospective study of 575 patients with chronic GVHD

Impairment criterion	Percent affected
Absolute increase in the sum of moveable and non-moveable sclerosis by >20%	4
Any increase in the global measure of skin sclerosis	12
Increase in the global skin score from 0 – 1 to 2 or 3, or from 2 to 3	9
Any increase in fascia score	10
Absolute increase in the sum of grade 3 – 4 skin involvement by >20%	3
Decrease of any joint range of motion score by ≥2 points	2
Increase of organ score from 0 – 2 to 3*	4
Oral ulceration	0
Esophagus	0
Upper gastrointestinal	0
Lower gastrointestinal	0
Eye	2
Joint or fascia	1
Genital	0
Lung	1
Increase of oral pain (page 4)score or GI global score from 0 – 1 to 2 – 3 or from 2 to 3	1
Increase of any other chronic GVHD indicator from 0 – 3 to 4	1
Decrease in Karnofsky Performance Score from 1 – 6 to 7 – 9	0
Decrease in the SF-36 physical function subscale of the SF-36 by ≥ 0.5 standard deviation	9
Decrease in HAP by ≥0.5 standard deviation	13
Decrease in SF-36 PCS by ≥0.5 standard deviation	8
Absolute decrease in FEV1 by ≥10% from first measurement	14

*includes oral ulceration, esophagus, upper GI, lower GI, eye, joint/fascia, genital, or lung score

In patients with progressive impairment, changes were measured between baseline and the first visit with progressive impairment. In the 130 patients without new impairment at any visit, changes were measured between baseline and the last available visit. Patients with progressive impairment had statistically significant worsening in clinician and patient-assessed chronic GVHD severity, chronic GVHD

symptom burden and quality of life as compared to the 130 patients who had no reported impairment at any follow-up visit (Table 11). Categorical comparisons based on criteria for improvement and worsening likewise showed statistically significant differences between patients with progressive impairment compared to those with no new impairment reported at any visit (Table 12). The results also showed that 2 – 8% of patients without impairment met criteria for worsening at the last follow-up visit by certain measures.

Table 11. Correlation of progressive impairment with changes in overall severity of chronic GVHD as assessed by providers and patients and by FACT-BMT and Lee Symptom Scale

Outcome measure	Change from enrollment to first visit with progressive impairment (N=237)*					Change from enrollment to last visit in patients without impairment (N=130)*					P
	N	Median	Mean	Min	Max	N	Median	Mean	Min	Max	
MD 0-3[†]	233	0	-0.15	-3	2	128	0	-0.59	-3	1	<0.001
MD 0-10[‡]	232	-1	-0.64	-8	6	128	-2	-1.98	-8	2	<0.001
PT 0-3[†]	179	0	-0.01	-2	2	59	-1	-0.58	-2	1	<0.001
PT 0-10[‡]	176	0	-0.37	-7	6	63	-2	-1.62	-6	3	<0.001
LSS	190	-0.27	-0.46	-45.22	33.57	63	-6.05	-7.59	-38.81	10.85	<0.001
FACT-BMT	180	-0.92	0	-44.67	54.67	60	13	15.00	-24.33	45.94	<0.001

*Negative values indicate improvement in all scales, except for the FACT-BMT

[†] Clinician (MD) or Patient (PT) -rated overall chronic GVHD severity on a 0-3 scale (none, mild, moderate, severe)

[‡] MD or PT-rated overall chronic GVHD severity on a 0-10 scale

LSS, Lee Symptom Scale; FACT-BMT, functional assessment of cancer therapy, bone marrow transplant

Table 12. Categorical changes in patients as assessed by providers and patients and by FACT-BMT and Lee Symptom Scale

Outcome measure	Change from enrollment to first visit with progressive impairment (N=237)			Change from enrollment to last visit in patients without impairment (N=130)			P
	Improved N (%)	No change N (%)	Worse N (%)	Improved N (%)	No change N (%)	Worse N (%)	
MD 0-3[*]	74 (32)	112 (48)	47 (20)	63 (49)	63 (49)	2 (2)	<0.001
MD 0-10[†]	82 (35)	112 (48)	38 (16)	72 (56)	53 (41)	3 (2)	<0.001
PT 0-3[*]	41 (23)	96 (55)	39 (22)	30 (51)	27 (46)	2 (3)	<0.001
PT 0-10[†]	46 (27)	87 (51)	38 (22)	34 (54)	24 (38)	5 (8)	<0.001
LSS	37 (19)	114 (60)	39 (21)	28 (44)	31 (49)	4 (6)	<0.001
FACT-BMT	27 (15)	122 (68)	31 (17)	36 (60)	22 (37)	2 (3)	<0.001

*Clinician (MD) or Patient (PT) -rated overall chronic GVHD severity on a 0-3 scale (none, mild, moderate, severe)

[†] MD or PT-rated overall chronic GVHD severity on a 0-10 scale

LSS, Lee Symptom Scale; FACT-BMT, functional assessment of cancer therapy, bone marrow transplant

In additional analyses, statistical and categorical changes in clinician and patient-assessed chronic GVHD severity, chronic GVHD symptom burden and quality of life were measured in patients who survived without progressive impairment (Tables 13 – 18). At 6, 12 and 24 months, patients without progressive impairment showed statistically significant improvement in all of these measures as compared to baseline. By these measures, 2% to 12% of the patients had worsening, with little difference in the proportions at 6, 12, and 24 months.

Table 13. Change scores in patients surviving at 6 months without progressive impairment (N = 176)

Measure	N	Mean	SD	Median	Min	Max	P*
MD 0-3	174	-0.43	0.74	0	-2	2	<0.001
MD 0-10	172	-1.44	1.90	-1	-7	4	<0.001
PT 0-3	111	-0.40	0.80	0	-3	2	<0.001
PT 0-10	113	-1.42	2.45	-1	-9	4	<0.001
Lee Symptom Scale	115	-5.85	9.05	-4.56	-41.02	30.99	<0.001
FACT-BMT	109	9.02	15.21	7.00	-32.57	70.00	<0.001

*signed rank test

Table 14. Categorical changes in patients surviving at 6 months without progressive impairment

Outcome Measure	Improved N (%)	No change N (%)	Worse N (%)	P*
MD 0-3†	69 (40)	96 (55)	9 (5)	<0.001
MD 0-10‡	76 (44)	88 (51)	8 (5)	<0.001
PT 0-3†	49 (45)	50 (45)	11 (10)	<0.001
PT 0-10‡	50 (45)	51 (46)	10 (9)	<0.001
Lee Symptom Scale	45 (39)	67 (58)	3 (3)	<0.001
FACT-BMT	38 (35)	65 (60)	6 (6)	<0.001

*binomial test for equal proportions between improved and worsened patients

Table 15. Change scores in patients surviving at 12 months without progressive impairment (N = 137)

Measure	N	Mean	SD	Median	Min	Max	P*
MD 0-3 change	134	-0.45	0.75	0	-2	1	<0.001
MD 0-10 change	132	-1.63	2.20	-1	-8	4	<0.001
PT 0-3 change	91	-0.35	0.71	0	-2	1	<0.001
PT 0-10 change	90	-1.43	2.39	-1.5	-7	3	<0.001
Lee Symptom Scale	96	-5.18	10.42	-3.97	-37.99	14.29	<0.001
FACT-BMT change	90	10.63	16.91	9.28	-38.00	58.72	<0.001

*signed rank test

Table 16. Categorical changes in patients surviving at 12 months without progressive impairment

Outcome Measure	Improved N (%)	No change N (%)	Worse N (%)	P*
MD 0-3†	57 (43)	68 (51)	9 (7)	<0.001
MD 0-10‡	65 (49)	60 (45)	7 (5)	<0.001
PT 0-3†	36 (40)	47 (52)	8 (9)	<0.001
PT 0-10‡	45 (51)	33 (37)	11 (12)	<0.001
Lee Symptom Scale	31 (32)	55 (57)	10 (10)	0.001
FACT-BMT	41 (46)	43 (48)	6 (7)	<0.001

*binomial test for equal proportions between improved and worsened patients

Table 17. Change scores in patients surviving at 24 months without progressive impairment (N = 53)

Measure	N	Mean	SD	Median	Min	Max	P*
MD 0-3	51	-0.47	0.73	0	-2	1	<0.001
MD 0-10	51	-1.69	2.05	-2	-6	2	<0.001
PT 0-3	31	-0.55	0.85	0	-2	1	0.002
PT 0-10	33	-1.70	2.47	-2	-7	4	<0.001
Lee Symptom Scale	30	12.05	13.47	10.67	-13	39	0.003
FACT-BMT	34	-5.45	10.16	-5.66	-39.23	11.45	<0.001

*signed rank test

Table 18. Categorical changes in patients surviving at 24 months without progressive impairment

Outcome Measure	Improved N (%)	No change N (%)	Worse N (%)	P*
MD 0-3[†]	23 (45)	25 (49)	3 (6)	<0.001
MD 0-10[‡]	28 (55)	22 (43)	1 (2)	<0.001
PT 0-3[†]	14 (45)	15 (48)	2 (6)	0.003
PT 0-10[‡]	17 (52)	14 (42)	2 (6)	0.001
Lee Symptom Scale	14 (41)	17 (50)	3 (9)	0.008
FACT-BMT	15 (50)	14 (47)	1 (3)	0.001

*binomial test for equal proportions between improved and worsened patients

Progressive impairment was not associated with a statistically significant increased risk of overall mortality (HR 0.85, 95% CI 0.59 – 1.22, p = 0.38) or mortality due to causes other than recurrent malignancy (HR 1.04, 95% CI 0.66 – 1.64, p = 0.84).

Discussion of survival without progressive impairment as an endpoint for chronic GVHD treatment trials

Prevention of "progressive impairment" equates to "survival without new impairment." This endpoint is conceptually similar to "progression-free survival" in oncology trials, in that "new impairment" captures the onset of any complication on a path leading toward an adverse effect on symptom burden and functional ability. The advantage of such an endpoint is that it could be applied at any stage in the course of the disease, with the expectation that a truly effective therapy for chronic GVHD should be able to prevent progression of major complications of the disease. Treatments that prevent unacceptable clinical deterioration in patients with chronic GVHD could be identified by using survival without progressive impairment as an endpoint in clinical trials. Therefore, this endpoint is highly relevant to the goals of treatment.

Methods for measuring progressive impairment have not been fully developed. In September, 2014, a questionnaire was sent to a group of clinicians with expertise in the management of chronic GVHD, asking whether they agreed that each change described in the definition represents progressive impairment and whether the change is likely due to chronic GVHD when it is observed in a patient

previously diagnosed with chronic GVHD. For most items, at least 15 of the 21 respondents agreed that each change represents progressive impairment. The following items were endorsed by smaller numbers of respondents.

Absolute increase in the sum of moveable and non-moveable sclerosis by >20%. This item was endorsed by 11 of the 21 respondents. At the same time, all respondents endorsed this item when the threshold was set at >50%. Taken together, these results indicate agreement that an increased body surface area affected by sclerosis represents progressive impairment, but confidence may be lacking that a 20% increase can be measured reliably or that this degree of change truly represents impairment.

Any increase in the measure of global skin sclerosis. This item was endorsed by 9 of the 21 respondents. Given the endorsement of other items related to skin sclerosis, confidence may be lacking that single-point score changes (i.e., 0 to 1, 1 to 2, 2 to 3, and 3 to 4) can be measured reliably. Ten of the 237 cases of progressive impairment were identified by this measure with no other indication of sclerosis or other evidence of new impairment identified by the provider.

Decrease in FEV1 by >10%. This item was endorsed by only 6 of the 21 respondents. The lack of endorsement almost certainly reflects the variability of FEV1 measurements and the potential confounding effects of infections and other causes of obstructive lung disease.

Decrease in the SF-36 physical component summary score by >0.5 standard deviation. This item was endorsed by only 6 of 19 respondents. At the same time, 19 of 20 respondents endorsed this item when the threshold was set at 1.0 standard deviation. Taken together, these results indicate agreement that a decrease in this measure reflects impairment, but confidence may be lacking that a 0.5 standard deviation change is clinically meaningful.

Decrease in the HAP by >0.5 standard deviation. This item was endorsed by only 6 of 19 respondents. At the same time, 19 of 20 respondents endorsed this item when the threshold was set at 1.0 standard deviation. Taken together, these results indicate agreement that a decrease in this measure reflects impairment, but confidence may be lacking that a 0.5 standard deviation change is clinically meaningful.

For most items, at least 16 of the 21 respondents agreed that each change is likely due to chronic GVHD. Only 9 of the respondents agreed that decreased FEV1 was likely due to chronic GVHD, and only 8 to 12 agreed that the patient-reported measures were likely due to chronic GVHD, regardless of whether the threshold was set at 0.5 or 1.0 standard deviation. In total, 44 cases of impairment were identified by FEV1 alone, and 70 were identified by patient measures with or without decreased FEV1. Therefore, 114 (48%) of the 237 cases of impairment were not substantiated by any provider measures. In the future, the specificity of using decreased FEV1 as an indicator of progressive impairment could be improved by asking providers for information that rules out causes other than chronic GVHD. With the current definition, survival without progressive impairment at 12 months is estimated to be 56%. Survival without progressive impairment at 12 months would increase to 74% if the definition included only the current provider-based measures.

Since methods for measuring progressive impairment have not been fully developed, benchmarks are not available for comparison with results of single-arm clinical trials. On the other hand, this endpoint could easily be applied in controlled trials for patients at any point in the course of the disease. Controlled two-arm trials with this endpoint could have more flexible eligibility criteria, since there is no need to match the characteristics of a benchmark cohort. Results would be informative for both efficacy and safety, leaving less uncertainty regarding the validity of comparisons inherent in single-arm trials.

The absence of non-relapse mortality and recurrent malignancy as components of survival without progressive impairment each reflect clinical benefit. Likewise, the absence of progressive impairment reflects clinical benefit, given that untreated "clinical extensive" chronic GVHD with a global NIH score of 2 or 3 leads inexorably to progressive impairment and eventually to major disability.

Survival without progressive impairment measures benefit as the absence of a new harm caused by the disease. Whereas the assessment of failure-free survival infers harm indirectly through treatment change, harm in the assessment of survival without progressive impairment is measured directly from a defined list of indicators. Further work is needed to establish agreement that each item in the list truly indicates reliably measured harm, that chronic GVHD is the most likely cause, and that important items have not been omitted from the list. With the current definition, the development of impairment was not associated with a statistically significant increased risk of overall mortality or mortality due to causes other than recurrent malignancy. Additional work is needed to determine whether this observation holds true with a refined definition of impairment.

As with failure-free survival, survival without progressive impairment does not directly measure improvement in manifestations of chronic GVHD. Paired comparisons from the prospective longitudinal observational study have suggested that patients who survive without progressive impairment at 6, 12 and 24 months have measurable overall reductions in symptom burden, disease activity and functional impairment, although the extent to which response shift might account for these results is not known. Results from the BMT CTN 0801 trial (NCT01106833) could determine whether these observations hold true in the context of a prospective clinical trial for chronic GVHD. In future controlled trials, secondary endpoints could compare the symptom burden and disease activity between arms when survival without progressive impairment is used as the primary endpoint. In future single arm trials, paired statistics could be used to determine whether patients surviving without progressive impairment have reductions in symptom burden and disease activity as compared to baseline.

Response

f. Are changes in clinician-reported chronic GVHD manifestations sufficient to document clinical benefit?

The 2005 NIH Consensus Conference proposed clinical response scales based on consensus. Multi-site measurement-development studies have evaluated reliability of the measures, and some of these studies and experience with the scoring algorithms since 2005 have helped to refine the scales. Assuming that data in the context of clinical trials can be accumulated to further validate measures in

the 2014 refined scoring algorithm and demonstrate correlation with clinician or patient perceptions of change, could the clinician-reported overall response measure serve as the primary endpoint in registration trials?"

Overall survival or survival to permanent resolution of chronic GVHD and discontinuation of systemic immunosuppression represent long-term clinical outcomes that are accepted as measures of meaningful benefit in chronic GVHD clinical trials, but these long-term outcomes are not practical for early-phase drug development studies. Qualitative assessments of chronic GVHD manifestations by transplant clinicians can guide clinical decisions and correlate with survival but are not adequate for reliable measurement of response in clinical trials. To accelerate development of novel therapeutic agents in chronic GVHD, quantitative research tools are needed for measuring short-term responses and predicting long-term clinical benefit.

The 2014 NIH Consensus Development Project on Criteria for Clinical Trials in chronic GVHD provides a firm framework of clinician-assessed core measures recommended for assessment of response in chronic GVHD systemic therapy trials (Appendix XI, Table 2, Form A). These measures include the following 13 elements all aimed to assess the most common chronic GVHD specific organ manifestations:

- NIH organ scores (skin, eye, joint-fascia, lung symptoms)
- Modified mouth score (OMRS)
- Total bilirubin and alkaline phosphatase
- FEV1 (% predicted)
- Photographic range of motion (P-ROM)
- GI scores (upper, lower and esophagus), and
- Physician assessed global rating scale

The semi-quantitative global assessment scales are included to detect qualitative improvements that are clinically meaningful but not always well captured using organ specific measures. Similar scales have been successfully incorporated in disease response algorithms in other immune-mediated diseases. Genital tract and other manifestations are not included due to lack of validated response measures or low frequency.

In contrast to the original 2005 NIH Consensus recommendations which were primarily expert driven,[28] this current set of measures is developed based on the accumulated evidence from prospective observational cohort studies and some phase II clinical intervention trials in chronic GVHD during the past decade. The 2014 NIH criteria also provide updated and practical recommendations for interpretation of organ specific responses and progression (Appendix XI, Table 4). Three general categories of overall response are proposed for interpretation of clinical trials:

- Complete response
- Partial response, and
- Lack of response (unchanged, mixed response, progression)

In Table 19 below we present details about the data supporting use of these NIH consensus recommended clinician-reported measures as documentation of clinical benefit in chronic GVHD intervention trials.

Table 19. Clinician-reported measures as potential indicators of benefit in clinical trials

Reference	Clinician - reported measures	Gold standard	Study design comments	Results
Mitchell[29]	Full 2005 NIH spectrum of measures – by transplant clinicians	Subspecialty experts	N=25 children and adults with chronic GVHD (4 consecutive pilot trials)	Supports feasibility of the NIH measures. Inter-rater agreement for skin and oral was satisfactory except for moveable sclerosis and moderate to substantial for functional capacity, GI and global rating measures.
Jacobsohn[30]	NIH skin score	Clinician and patient perception of skin improvement or worsening, Overall survival	N=458 prospective multicenter longitudinal observational cohort study	The 0-3 NIH composite skin score correlated with both clinician and patient perception of improvement or worsening. Worsening skin score at 6 months was associated with worse survival.
Inamoto[31]	NIH eye score	Clinician and patient perception of eye symptom change	N=387 prospective multicenter longitudinal observational cohort study	Among all scales, changes in the NIH eye scores showed the greatest sensitivity to symptom change reported by clinicians or patients. Schirmer's test did not correlate.
Treister[32]	NIH oral score and modified OMRS (0-15)	Patient and clinician-reported change in oral chronic GVHD	N=458 prospective multicenter longitudinal observational cohort study	The clinician-reported measurement changes most predictive of perceived change by clinicians and patients were erythema, extent of lichenoid changes, and NIH severity score.
Palmer[33]	NIH lung score symptom scale	Non-relapse mortality (NRM), Overall survival (OS), Patient-reported lung symptoms	N=496 prospective multicenter longitudinal observational cohort study	The NIH symptom-based lung score was associated with NRM, OS, patient-reported symptoms, and functional status. Worsening of NIH symptom-based lung score over time was associated with higher NRM and lower survival.
Inamoto[34]	NIH joint-fascia score, Hopkins scale, Photographic (P-ROM)	Clinician and patient perception of change	N=567 prospective multicenter longitudinal observational cohort study	Changes in the NIH scale correlated with both clinician- and patient-perceived improvement. Changes in all 3 scales correlated with clinician- and patient-perceived worsening, but the P-ROM scale was the most sensitive.

Bassim[35]	NIH modified OMRS (0-15)	Established measures of oral pain, oral function, oral related QOL, nutrition and laboratory parameters.	N=198 prospective cross-sectional observational cohort study (moderate-to-severe chronic GVHD)	This study supports the use of the OMRS and its components (erythema, lichenoid and ulcerations) to measure clinician-reported severity of oral chronic GVHD. No associations were found between mucoceles and any indicator evaluated.
Curtis[36]	18 clinician-reported ('Form A') measures	Concurrent parameters: NIH global score, chronic GVHD activity, Lee symptom score and SF36 PCS	N=193 prospective cross-sectional observational cohort study (moderate-to-severe chronic GVHD)	4-point and 11-point clinician reported global symptom severity scores are associated with the majority of concurrent outcomes. Skin erythema is a potentially reversible sign of chronic GVHD that is associated with survival.
Yanik[37]	Response was defined as 10% FEV1 or FVC improvement	5-year survival	N=34 patients with subacute pulmonary dysfunction (25 obstructive) received etanercept therapy	5-year survival 90% (95% CI, 73%-100%) for 10 patients who responded to therapy, compared with 55% (95% CI, 37%-83%) for the 21 patients who did not meet response criteria (P = 0.07)
Olivieri[38]	NIH criteria, NIH organ score, Couriel criteria	Overall survival	N=40, Phase II prospective study of imatinib for steroid-refractory chronic GVHD	The 3-year OS was 94% for patients responding at 6 months and 58% for non-responders according to NIH response criteria (P = 0.007)
BMT CTN 0801 (unpublished)	NIH criteria	Clinician assessed overall CR+PR	N=105, randomized phase II multicenter trial	AUC for organs (lichenoid mouth, joint score) plus clinician assessed 0-10 global rating scale = 0.79

g. Could improvement in a patient-reported outcome (PRO) tool be considered sufficient documentation of clinical benefit?

Incorporation of the patient experience into endpoints for clinical trials addresses the "living better" portion of the definition of "clinical benefit." For a disease such as chronic GVHD, quality of life and symptoms may reflect disease activity, residual effects of GVHD or the side effects of medications used to treat GVHD. FDA has released draft guidance for qualification of PRO instruments.[39] This guidance outlines steps necessary to consider a PRO instrument adequate to measure clinical benefit for purposes of regulatory approval.

Growing evidence supports the validity of patient-reported instruments. The Lee Symptom Scale is a 30-item, 7-domain symptom scale that has proven reliable, valid, and sensitive to change. It was developed with patient input and tested in a cohort of 107 patients with active chronic GVHD who completed the questionnaire every 3 or 6 months. Psychometric properties have been published.[23] Subsequent studies have shown that changes in the eye, skin, mouth, GI, and summary scale have correlated with patient- and clinician-reported changes in chronic GVHD activity.[31,32,40,41] However, this instrument is not

sufficient for use as a primary endpoint in clinical trials making general claims, even if it were qualified by the FDA, because it measures only symptoms, limiting labeling claims to symptom improvement. In addition, although most symptoms are specific to chronic GVHD activity, the interpretation of changes may be confounded by adverse side effects of treatment. Finally, most trials of chronic GVHD treatment are not blinded, raising concerns about the validity of PROs that can be affected by patient beliefs that an active drug is being administered.

The only other chronic GVHD-specific scale is the MD Anderson chronic GVHD symptom scale, published only in abstract form, and modeled after the MD Anderson Symptom Inventory (MDASI). Almost no work in chronic GVHD has used the Patient Reported Outcomes Measurement Information System (PROMIS) instruments.

Multi-dimensional health-related quality of life (HR-QOL) instruments such as the MOS SF-36 (Medical Outcomes Study Short Form 36)[42,43] and the FACT-BMT (Functional assessment of cancer therapy – bone marrow transplantation subscale)[21,44] have been used in many trials. In general, these instruments are able to detect differences according to the occurrence of chronic GVHD,[45] severity of chronic GVHD[46] and change in chronic GVHD activity as reported by patients and clinicians,[47] but not when compared with NIH calculated responses.[41] Many multi-domain HR-QOL instruments lack sensitivity to changes in specific syndromes associated with disease states. Alternatively, it is possible that NIH-calculated response measures do not accurately capture changes of value or importance to patients and clinicians.

Table 20 provides details about studies addressing PROs sensitivity to change, the most difficult criterion to demonstrate for an instrument.

Table 20. PROs sensitivity to change

Reference	Patient-reported measure	Gold standard	Study design comments	Results
Global				
Pidala[47]	SF-36, FACT-BMT	Change in global severity, clinician-reported, patient-reported change	N=336, correlation of change scores with response measures in an observational study	Patient-reported severity change was associated with all QOL measures. Change in NIH and clinician-reported chronic GVHD severity did not correlate well with patient-reported QOL changes.
Inamoto[41]	SF-36, FACT-BMT, Lee symptom scale	NIH-calculated overall response	N=258, correlation of change scores with NIH-calculated overall response in an observational study	NIH calculated overall responses were associated with patient-reported symptoms in patients enrolled within 3 months of chronic GVHD onset but not in patients enrolled more than 3 months after onset. SF-36 and FACT-BMT changes were not associated with NIH-calculated responses regardless of time since onset.

Walker[48]	SF-36, FACT-BMT, Lee symptom scale	N/A	N=203, randomized, unblinded study of thymoglobulin vs. no thymoglobulin, comparing PROs between randomized groups	The study met its primary endpoint: freedom from immunosuppressive treatment at 12 months (37.4% vs. 16.5%, P = 0.001). GVHD symptoms were lower in patients randomized to Thymoglobulin (14.95 vs. 20.93, P = 0.017). The difference was also clinically meaningful, defined via the distribution method as 0.5 SD.
Organ-specific				
Inamoto[31]	0-10 eye symptom, Lee eye symptom score, ocular surface disease index (OSDI)	Patient and clinician-reported change in eye chronic GVHD (8-point scale)	N=387, correlation of PRO change scores with reported response in an observational study	Change in the Lee eye symptom score, 0-10 eye symptom, and OSDI correlated with patient- and clinician-reported change
Jacobsohn[30]	Lee skin symptom score	Non-relapse mortality, overall survival, patient- and clinician-reported change (8-point scale)	N=458, correlation with outcomes and reported change in an observational study	Change in the Lee skin symptom score correlated with patient and clinician-perceived changes. Improvement in the Lee skin symptoms score at 6 months was associated with lower NRM and better OS
Treister[32]	Lee mouth and nutrition symptom scores, patient mouth sensitivity, pain, dryness 0-10	Patient- and clinician-reported change in oral chronic GVHD (8-point scale)	N=458, correlation with reported change in an observational study	In multivariate modeling, change in patient-reported Lee mouth symptom score was associated with patient- and clinician-reported change
Inamoto[34]	Lee muscle/joint symptom score, global GVHD severity 0-10, SF-36, FACT-BMT	Patient- and clinician-reported change in joint chronic GVHD (8-point scale)	N=567, correlation with reported change in an observational study	Change in the Lee muscle/joint symptom score, overall symptom score and 0-10 global score correlated with patient-reported improvement and worsening of joint GVHD and clinician-reported worsening of joint GVHD. SF-36 PCS correlated with patient- and clinician-reported improvement in joint GVHD. FACT-G correlated with patient- and clinician-reported worsening in joint GVHD.
Inamoto[41]	Lee symptom scale, mouth, eye, skin 0-10 symptoms	NIH-calculated organ-specific change	N=258, correlation with NIH-calculated organ changes in an observational study	NIH calculated organ responses were associated with patient-reported symptom change in skin, eye, mouth and GI (nutrition).

SF-36, Medical Outcomes Study Short Form-36; FACT-BMT, Functional Assessment of Cancer Therapy, Bone Marrow Transplantation subscale; NRM, non-relapse mortality; OS, overall survival

The NIH Consensus conference recommended PROs as secondary endpoints in the 2005 Clinical Trials Working Group Report, primarily because of challenges inherent in the collection and interpretation of PROs, and lack of a qualified instrument. The field may be interested in pursuing qualification of PRO instruments if FDA were willing to accept PROs as a primary endpoint in chronic GVHD studies. Alternatively, PROs could serve as key secondary endpoints, or components of PROs could be incorporated into a composite global scale.

h. Would changes in a multicomponent clinical scale that incorporates clinician assessments, patient-reported outcomes and laboratory or functional measurements be sufficient to document changes in chronic GVHD disease activity?

Validated scales that incorporate clinician assessments (e.g., on a 0-10 or global scale, or organ measures), patient-reported outcomes (e.g., symptoms or quality of life), and laboratory or functional measures (e.g., C-reactive protein) have been used as the primary endpoints in registration trials for other immune-mediated diseases such as lupus,[49-55] Crohn's disease,[56,57] ankylosing spondylitis,[58] and rheumatoid arthritis.[59,60] These scales were generally developed by identifying clinical, laboratory and patient-reported parameters associated with reported perceptions of change or changes in management (e.g., adding or decreasing immunosuppressive treatment). Table 21 outlines the factors included in multi-component clinical scales in other diseases.

No such aggregate scale exists for chronic GVHD. In a preliminary analysis, the Chronic GVHD Consortium used observational data from 497 patients accrued from ten sites to identify changes in chronic GVHD-associated variables that are independently correlated with clinician-reported complete or partial response after 6 months. First, 12 separate multivariable models that included all measures of each organ or concept were evaluated (skin – 58, mouth – 10, eye – 8, GI – 7, liver – 5, joint – 6, lung – 15, patient-reported outcomes – 15, functioning – 6, overall – 6, laboratory – 14, other – 14). Changes in factors that were significant at $p<0.05$ were included in a single multivariable model to identify independent correlates at $p<0.01$. The first multi-dimension model did not include clinician-reported 0-10 severity because of concern that this scale might be viewed as less objective than individual organ measures. Five items were identified as correlated with clinician-reported complete or partial response at 6 months: Vienna skin scale, lichenoid mouth changes, NIH eye score, NIH joint score, patient self-reported nutrition symptoms. When clinician-reported overall severity 0-10 was allowed into the model, three factors were identified as independently correlated: change in clinician 0-10 overall chronic GVHD severity, change in clinician-reported NIH joint score 0-3, and change in extent of oral lichenoid changes (0-3). This model had an AUC of 0.82 in the training set (Figure 7A). The model was then applied directly to 105 participants in Blood and Marrow Transplant Clinical Trials Network (BMT CTN) 0801. Although BMT CTN was a randomized therapeutic study, all results were analyzed with blinding to the study intervention using the final 3-factor model. The AUC of the 3-factor model was 0.79 for association with clinician-reported overall complete or partial response at 6 months (Figure 7B).

Table 21. Components of clinical scales in other diseases

	Systemic lupus	Ankylosing spondylitis	Rheumatoid arthritis	Crohn's disease
Measure	Responder Index	ASA20	ACR20	CDAI
Patient-reported		• Patient global • Pain • Physical function • Morning stiffness	• Patient global • Pain • Physical function	• Number of liquid or soft stools • Abdominal pain • General well-being
Physician-reported	• Physician global • SLEDAI (composite of signs/labs, 0-108) • BILAG (activity in each of 9 organ systems, based on treatment required)		• Physician global • Swollen joint count	• Number of complications • Anti-diarrheal agents • Abdominal mass
Laboratory			• ESR or CRP	• Hematocrit • Body weight
Criteria for improvement	• ≥ 4 point improvement in SLEDAI • No new/worsened organ systems in BILAG • No worsening in physician global	• ≥ 20% improvement (and ≥ 1 point on 0-10 scale) in 3 measures • No worsening in 4th	• ≥ 20% improvement in both joint counts • ≥20% improvement in 3 other measures	• Induction of remission, CDAI < 150 • Decrease in CDAI >50-100

ASA, Ankylosing Spondylitis Assessment; ACR, American College of Rheumatology; CDAI, Crohn's Disease Activity Index; SLEDI, Systemic Lupus Erythematosus Disease Activity Index; BILAG, British Isles Lupus Assessment Group; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein

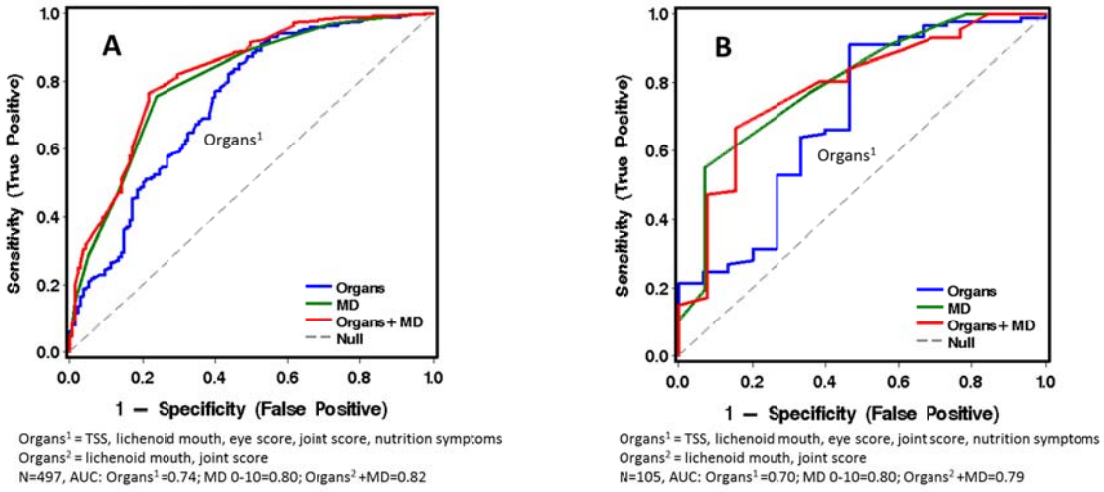


Figure 7. Receiver operator curve showing performance of a 3-factor model as a correlate of clinician-reported overall complete or partial response at 6 months. A) Training cohort, B) Verification cohort. Factors include change in clinician 0-10 overall chronic GVHD severity, change in clinician-reported NIH joint score 0-3, and change in extent of oral lichenoid changes (0-3).

The three items that comprise the model are shown below.

Where would you rate the severity of this patient's chronic GvHD symptoms on the following scale, where 0 is cGVHD symptoms that are not at all severe and 10 is the most severe cGVHD symptoms possible:

0 1 2 3 4 5 6 7 8 9 10

cGvHD symptoms not at all severe — **Most severe cGvHD symptoms possible**

JOINTS AND FASCIA	☐ No symptoms	☐ Mild tightness of arms or legs, normal or mild decreased range of motion (ROM) AND not affecting ADL	☐ Tightness of arms or legs OR joint contractures, erythema thought due to fasciitis, moderate decrease ROM AND mild to moderate limitation of ADL	☐ Contractures WITH significant decrease of ROM AND significant limitation of ADL (unable to tie shoes, button shirts, dress self etc.)

Lichenoid	None	0	Lichen-like changes (<25%)	1	Lichen-like changes (25-50%)	2	Lichen-like changes (>50%)	3

The value of including a variety of measures reflecting different aspects of a disease process is codified in the Outcomes Measures in Rheumatology (OMERACT) effort.[61,62] The OMERACT consensus initiative specifies the process of identifying a core set of measures that should be included in any randomized controlled trial or long-term observational study in a rheumatologic disease, including incorporation of the patient perspective from the start of the process. The framework includes four areas: Death, Life Impact, Resource Utilization, and Pathophysiological Manifestations. Life impact is generally assessed by PROs. Pathophysiologic manifestations are measured by physical exam or laboratory testing. The OMERACT filter requires that one measure in each area be identified as a core measure.

i. If yes, then what gold standard of clinical benefit should be used to develop and validate such a scale?

If the chronic GVHD research community were to follow the example in rheumatology and develop a composite scale drawn from clinician-report, patient-reported and objectively measured data, then the gold-standard anchor must be identified. In the chronic GVHD example above, clinician-reported complete and partial response at 6 months was used as the gold standard, since this surrogate endpoint correlates with subsequent NRM and OS.[63] Other potential anchors could be patient-reported overall response or change in chronic GVHD disease activity. Anchors could be reported by patients or clinicians as change on a Likert or 0-10 scale or could be inferred from medication management such as addition of a new therapy or tapering and discontinuation of a current therapy (Figure 8). Survival, non-relapse mortality, NIH calculated response, failure-free survival and survival without progressive impairment are other possible anchors.

We favor using some direct measure of change in chronic GVHD activity as the anchor for this exercise, such as the 8 point assessment scale in Figure 8 completed by clinicians or patients, since this is the most proximal measure of change in chronic GVHD experienced by the patient. The use of more distal measures such as non-relapse mortality or failure-free survival increase the risk that factors unrelated to GVHD will influence the anchor.

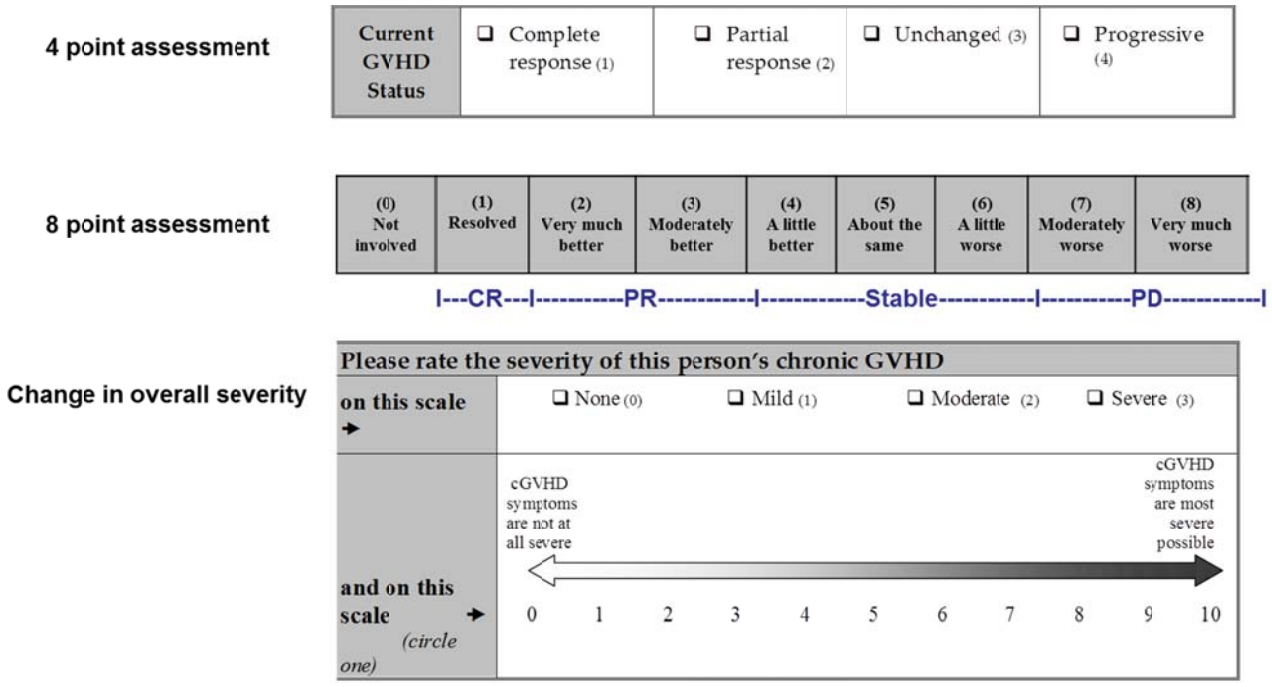


Figure 8. Potential patient- and clinician-reported measures of change in chronic GVHD activity.

References

1. CIBMTR: http/www.cibmtr.org/ReferenceCenter/SlidesReports/SummarySlides/pages/index.aspx.
2. Flowers ME, Inamoto Y, Carpenter PA, et al. Comparative analysis of risk factors for acute graft-versus-host disease and for chronic graft-versus-host disease according to National Institutes of Health consensus criteria. *Blood*. 2011;117(11):3214-3219.
3. Vigorito AC, Campregher PV, Storer BE, et al. Evaluation of NIH consensus criteria for classification of late acute and chronic GVHD. *Blood*. 2009;114(3):702-708.
4. Arora M, Wagner JE, Davies SM, et al. Randomized clinical trial of thalidomide, cyclosporine, and prednisone versus cyclosporine and prednisone as initial therapy for chronic graft- versus-host disease. *Biol Blood Marrow Transplant*. 2001;7(5):265-273.
5. Gilman AL, Schultz KR, Goldman FD, et al. Randomized trial of hydroxychloroquine for newly diagnosed chronic graft-versus-host disease in children: a Children's Oncology Group study. *Biol Blood Marrow Transplant*. 2012;18(1):84-91.
6. Koc S, Leisenring W, Flowers ME, et al. Thalidomide for treatment of patients with chronic graft-versus-host disease. *Blood*. 2000;96(12):3995-3996.
7. Koc S, Leisenring W, Flowers MED, et al. Therapy for chronic graft-versus-host disease: a randomized trial comparing cyclosporine plus prednisone versus prednisone alone. *Blood*. 2002;100:48-51.
8. 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.
9. Sullivan KM, Witherspoon RP, Storb R, et al. Prednisone and azathioprine compared with prednisone and placebo for treatment of chronic graft-v-host disease: prognostic influence of prolonged thrombocytopenia after allogeneic marrow transplantation. *Blood*. 1988;72(2):546-554.
10. Flowers ME, Apperley JF, van Besien K, et al. A multicenter prospective phase 2 randomized study of extracorporeal photopheresis for treatment of chronic graft-versus-host disease. *Blood*. 2008;112(7):2667-2674.
11. Greinix HT, van Besien K, Elmaagacli AH, et al. Progressive improvement in cutaneous and extracutaneous chronic graft-versus-host disease after a 24-week course of extracorporeal photopheresis--results of a crossover randomized study. *Biol Blood Marrow Transplant*. 2011;17(12):1775-1782.
12. Martin PJ, Inamoto Y, Carpenter PA, Lee SJ, Flowers ME. Treatment of chronic graft-versus-host disease: Past, present and future. *Korean J Hematol*. 2011;46(3):153-163.
13. 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.
14. Wolff D, Gerbitz A, Ayuk F, et al. Consensus conference on clinical practice in chronic graft-versus-host disease (GVHD): first-line and topical treatment of chronic GVHD. *Biol Blood Marrow Transplant*. 2010;16(12):1611-1628.
15. Wolff D, Schleuning M, von Harsdorf S, et al. Consensus Conference on Clinical Practice in Chronic GVHD: Second-Line Treatment of Chronic Graft-versus-Host Disease. *Biol Blood Marrow Transplant*. 2011;17(1):1-17.
16. Martin PJ, Weisdorf D, Przepiorka D, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: VI. Design of Clinical Trials Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(5):491-505.
17. Inamoto Y, Flowers ME, Sandmaier BM, et al. Failure-free survival after initial systemic treatment of chronic graft-versus-host disease. *Blood*. 2014;124(8):1363-1371.

18. Chronic GVHD Consortium. Rationale and design of the chronic GVHD cohort study: improving outcomes assessment in chronic GVHD. *Biol Blood Marrow Transplant*. 2011;17(8):1114-1120.
19. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes*. 2003;1:4.
20. Ware JE, Jr. SF-36 health survey update. *Spine (Phila Pa 1976)*. 2000;25(24):3130-3139.
21. McQuellon RP, Russell GB, Cella DF, et al. Quality of life measurement in bone marrow transplantation: development of the Functional Assessment of Cancer Therapy-Bone Marrow Transplant (FACT-BMT) scale. *Bone Marrow Transplant*. 1997;19(4):357-368.
22. Daughton DM, Fix AJ, Kass I, Bell CW, Patil KD. Maximum oxygen consumption and the ADAPT quality-of-life scale. *Arch Phys Med Rehabil*. 1982;63(12):620-622.
23. Lee S, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2002;8(8):444-452.
24. Lee SJ, Wolff D, Kitko C, et al. Measuring therapeutic response in chronic graft-versus-host disease. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-Versus-Host Disease: IV. The 2014 Response Criteria Working Group Report. Biol Blood Marrow Transplant, In preparation.
25. Inamoto Y, Storer BE, Lee SJ, et al. Failure-free survival after second-line systemic treatment of chronic graft-versus-host disease. *Blood*. 2013;121(12):2340-2346.
26. Jacobsohn DA. FFS: an end(point) to our problems in chronic GVHD trials? *Blood*. 2014;124(8):1216-1217.
27. Vogelsang GB. Order out of chaos. *Blood*. 2013;121(12):2170-2172.
28. Pavletic SZ, Martin P, Lee SJ, et al. Measuring Therapeutic Response in Chronic Graft-versus-Host Disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: IV. Response Criteria Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(3):252-266.
29. Mitchell SA, Jacobsohn D, Thormann Powers KE, et al. A Multicenter Pilot Evaluation of the National Institutes of Health Chronic Graft-versus-Host Disease (cGVHD) Therapeutic Response Measures: Feasibility, Interrater Reliability, and Minimum Detectable Change. *Biol Blood Marrow Transplant*. 2011.
30. Jacobsohn DA, Kurland BF, Pidala J, et al. Correlation between NIH composite skin score, patient-reported skin score, and outcome: results from the Chronic GVHD Consortium. *Blood*. 2012;120(13):2545-2552; quiz 2774.
31. Inamoto Y, Chai X, Kurland BF, et al. Validation of measurement scales in ocular graft-versus-host disease. *Ophthalmology*. 2012;119(3):487-493.
32. Treister N, Chai X, Kurland B, et al. Measurement of oral chronic GVHD: results from the Chronic GVHD Consortium. *Bone Marrow Transplant*. 2013.
33. Palmer J, Williams K, Inamoto Y, et al. Pulmonary symptoms measured by the national institutes of health lung score predict overall survival, nonrelapse mortality, and patient-reported outcomes in chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2014;20(3):337-344.
34. Inamoto Y, Pidala J, Chai X, et al. Assessment of joint and fascia manifestations in chronic graft-versus-host disease. *Arthritis Rheumatol*. 2014;66(4):1044-1052.
35. Bassim CW, Fassil H, Mays JW, et al. Validation of the National Institutes of Health chronic GVHD Oral Mucosal Score using component-specific measures. *Bone Marrow Transplant*. 2014;49(1):116-121.
36. Curtis LM, Grkovic L, Mitchell SA, et al. NIH response criteria measures are associated with important parameters of disease severity in patients with chronic GVHD. *Bone Marrow Transplant*. 2014.

37. Yanik GA, Mineishi S, Levine JE, et al. Soluble tumor necrosis factor receptor: enbrel (etanercept) for subacute pulmonary dysfunction following allogeneic stem cell transplantation. *Biol Blood Marrow Transplant*. 2012;18(7):1044-1054.
38. Olivieri A, Cimminiello M, Corradini P, et al. Long-term outcome and prospective validation of NIH response criteria in 39 patients receiving imatinib for steroid-refractory chronic GVHD. *Blood*. 2013.
39. Health USDo, Human Services FDACfDE, Research, et al. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4:79.
40. Jacobsohn DA, Rademaker A, Kaup M, Vogelsang GB. Skin response using NIH consensus criteria vs Hopkins scale in a phase II study for steroid-refractory chronic GVHD. *Bone Marrow Transplant*. 2009;44(12):813-819.
41. Inamoto Y, Martin PJ, Chai X, et al. Clinical benefit of response in chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2012;18(10):1517-1524.
42. Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston: The Health Institute, New England Medical Center; 1994.
43. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey: a manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.
44. McQuellon RP, Russell GB, Rambo TD, et al. Quality of life and psychological distress of bone marrow transplant recipients: the 'time trajectory' to recovery over the first year. *Bone Marrow Transplant*. 1998;21(5):477-486.
45. Lee SJ, Kim HT, Ho VT, et al. Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant*. 2006;38(4):305-310.
46. Pidala J, Kurland B, Chai X, et al. Patient-reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. *Blood*. 2011;117(17):4651-4657.
47. Pidala J, Kurland BF, Chai X, et al. Sensitivity of changes in chronic graft-versus-host disease activity to changes in patient-reported quality of life: results from the Chronic Graft-versus-Host Disease Consortium. *Haematologica*. 2011;96(10):1528-1535.
48. Walker I, Schultz KR, Toze CL, et al. Thymoglobulin decreases the need for immunosuppression at 12 monhs after myeloablative and nonmyeloablative unrelated donor transplantation: CBMTG 0801, a randomized, controlled trial. 2014 American Society of Hematology Abstract.
49. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. *Arthritis Rheum*. 1992;35(6):630-640.
50. Furie R, Petri M, Zamani O, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum*. 2011;63(12):3918-3930.
51. Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*. 2002;29(2):288-291.
52. Hay EM, Bacon PA, Gordon C, et al. The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus. *Q J Med*. 1993;86(7):447-458.
53. Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. *Arthritis Rheum*. 1989;32(9):1107-1118.
54. Luijten KM, Tekstra J, Bijlsma JW, Bijl M. The Systemic Lupus Erythematosus Responder Index (SRI); a new SLE disease activity assessment. *Autoimmun Rev*. 2012;11(5):326-329.

55. Vitali C, Bencivelli W, Isenberg DA, et al. Disease activity in systemic lupus erythematosus: report of the Consensus Study Group of the European Workshop for Rheumatology Research. II. Identification of the variables indicative of disease activity and their use in the development of an activity score. The European Consensus Study Group for Disease Activity in SLE. *Clin Exp Rheumatol*. 1992;10(5):541-547.
56. Best WR, Becktel JM, Singleton JW, Kern F, Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology*. 1976;70(3):439-444.
57. Sandborn WJ, Feagan BG, Hanauer SB, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. *Gastroenterology*. 2002;122(2):512-530.
58. Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. *Arthritis Rheum*. 2001;44(8):1876-1886.
59. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum*. 1995;38(6):727-735.
60. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis*. 2011;70(3):404-413.
61. Boers M, Idzerda L, Kirwan JR, et al. Toward a generalized framework of core measurement areas in clinical trials: a position paper for OMERACT 11. *J Rheumatol*. 2014;41(5):978-985.
62. Kirwan JR, Boers M, Tugwell P. Updating the OMERACT filter at OMERACT 11. *J Rheumatol*. 2014;41(5):975-977.
63. Palmer J, Chai X, Martin PJ, et al. Physician-reported CR+PR at 6 months predicts subsequent survival in patients wtih chronic GVHD. 2015 BMT Tandem meeting abstract.

Supplementary Tables

Supplementary Table 1. Randomized trials of treatment for chronic GVHD

Author	Arms compared	Double-blind	N	Results
Sullivan[9]	Initial prednisone ± azathioprine	Yes	179	Decreased survival
Koc[7]	Initial prednisone ± cyclosporine	No	287	Possible steroid-sparing effect
Koc[6]	Initial CSP/prednisone ± thalidomide	Yes	51	Toxicity
Arora[4]	Initial CSP/prednisone ± thalidomide	No	54	No benefit
Martin[8]	Initial CNI/prednisone ± MMF	Yes	151	No benefit
Gilman[5]	Initial CNI/prednisone ± hydroxychloroquine	No	54	Terminated early
Flowers[10] Greinix[11]	Second-line extracorporeal photopheresis	Yes	95	Possible steroid-sparing effect
Carpenter*	Initial sirolimus/prednisone ± CNI	No	151	Terminated early

*results not yet reported

Supplementary Table 2. Change scores in patients with failure-free survival at 1 year after enrollment

Variable	N	Mean*	SD	Median	Min	Max	P†
Physician 0-3 chronic GVHD severity score	67	-0.63	0.81	-1	-2	1	<.001
Physician 0-10 chronic GVHD severity score	67	-2.12	2.09	-2	-6	3	<.001
Patient 0-10 rating of overall GVHD symptoms	42	-1.45	2.38	-1.5	-6	4	<.001
Patient 0-3 overall severity	47	-0.36	0.76	0	-2	1	0.001
Lee Symptom Scale skin	47	-9.87	13.85	-6.25	-45	10	<.001
Lee Symptom Scale energy	45	-4.24	15.24	0	-50	22.02	0.07
Lee Symptom Scale lung	46	**2.07**	14.97	0	-20	60	0.8
Lee Symptom Scale eye	46	-0.91	24.33	-6.25	-50	58.33	0.45
Lee Symptom Scale nutrition	45	-3.67	10.57	0	-40	25	0.02
Lee Symptom Scale psychological	44	-5.97	18.45	0	-75	29.17	0.02
Lee Symptom Scale mouth	47	-4.52	26.25	0	-75	62.5	0.22
Lee Symptom Scale summary	47	-3.39	9.64	-4.06	-28.45	21.48	0.004
FACT physical well-being	48	1.49	4.42	1	-10	19	0.007
FACT social/family well-being	48	0.97	3.39	1	-7	12.83	0.04
FACT emotional well-being	47	1.19	3.40	1	-5	16	0.03
FACT functional well-being	47	4.23	5.16	3.5	-5	19	<.001
FACT BMT subscale	47	1.94	5.07	2	-8	22	0.009
FACT-BMT trial outcome index	47	7.90	11.88	7.22	-13.44	60	<.001
FACT total (FACT-G)	46	8.16	12.18	7.5	-13.5	66.83	<.001
FACT-BMT total score	46	10.21	16.37	10	-17.94	88.83	<.001
Karnofsy Performance Score	48	4.38	13.03	0	-50	30	0.003
SF36 physical functioning Scale	48	2.25	7.14	2.81	-29.46	16.84	0.002
SF36 Role-Functioning Physical Scale	48	6.80	9.64	6.12	-19.59	24.49	<.001
SF36 Bodily Pain Scale	48	0.10	8.99	0	-20.71	29.16	0.87
SF36 General Health Perceptions Scale	48	2.31	10.44	1.19	-17.64	36.7	0.18
SF36 Vitality Scale	48	2.25	8.81	1.56	-12.49	28.1	0.22
SF36 Social Functioning Scale	48	5.57	12.85	5.45	-21.82	38.18	0.006
SF36 Role-Functioning Emotional Scale	48	4.13	10.12	0	-11.66	38.87	0.005
SF36 Mental Health Scale	48	0.29	10.05	0	-30.98	42.24	0.98
SF36 Physical Component Scale	48	3.12	7.18	1.63	-12.18	20.2	0.007
SF36 Mental Component Scale	48	2.56	10.28	1.07	-25.02	44.34	0.12
Human Activity Profile maximum activity score	48	3.29	10.43	0	-24	28	0.04
Human Activity Profile adjusted activity score	48	6.10	12.93	5	-37	39	<.001
Modified Human Activity Profile adjusted activity score	48	4.21	11.36	4	-35	31	0.003
NIH 0-3 global severity	67	-0.37	0.83	0	-3	1	<.001

*****Worsened measures are highlighted in bold (not statistically significant).
† signed rank test

Supplementary Table 3. Change scores in patients with failure-free survival at 1 year after enrollment

Variable	N*	Mean†	SD	Median	Min	Max	P‡
Body surface area erythema	35	-30.40	28.50	-21.6	-100	3.6	<.001
Body surface area movable sclerosis	12	-0.83	9.59	2.7	-19.8	9	0.83
Body surface area non-movable sclerosis	6	-10.95	17.99	-9	-45	4.5	0.13
Skin sclerotic score (0-4)	12	-0.42	1.68	0	-4	1	0.68
NIH skin score (0-3)	40	-0.98	1.27	-1	-3	2	<.001
Fascia score (0-3)	12	-0.50	1.31	-1	-3	1	0.3
Vienna skin score summary measure (0-100)	45	-2.31	3.46	-1.4	-13.7	4.5	<.001
Vienna skin score sum of grade 3 and 4	11	-0.64	9.05	2.1	-21.6	9	0.92
Range of motion shoulder score (1-7)	6	0.17	0.98	0.5	-1	1	0.99
Range of motion elbow score (1-7)	6	0.17	2.56	-1	-2	5	0.78
Range of motion wrist score (1-7)	15	**-0.33**	0.98	-1	-2	1	0.31
Range of motion foot score (1-4)	17	**-0.35**	0.70	0	-1	1	0.11
NIH mouth score (0-3)	52	-0.63	0.91	-1	-2	2	<.001
Mouth erythema score (0-3)	34	-0.65	1.28	-1	-3	3	0.003
Mouth lichenoid score (0-3)	53	-0.77	1.10	-1	-3	2	<.001
Mouth ulcers score (0-3)	13	-0.62	1.89	-2	-2	2	0.39
Mouth mucoceles score (0-3)	28	-0.54	1.73	-1	-3	3	0.1
Mouth pain score (0-3)	34	-0.53	0.75	-1	-2	1	<.001
NIH GI score (0-3)	24	-0.92	0.97	-1	-3	1	<.001
GI esophagus score (0-3)	9	-1.11	1.27	-1	-3	1	0.06
Upper GI score (0-3)	16	-1.38	1.02	-1	-3	1	<.001
Lower GI score (0-3)	10	-0.60	1.35	-1	-3	1	0.29
NIH eye score (0-3)	39	-0.08	0.96	0	-2	2	0.63
NIH joint score (0-3)	24	-0.13	1.19	0	-2	2	0.54
NIH genital score (0-3)	10	**0.20**	1.48	0.5	-2	2	0.83
Physician reported lung symptom score (0-3)	19	-0.26	1.15	-1	-2	1	0.32
Patient 0-10 rating of skin itching	35	-1.29	2.55	-1	-6	3	0.007
Patient 0-10 rating of mouth dryness	38	-2.42	3.01	-2.5	-7	7	<.001
Patient 0-10 rating of mouth pain	27	-1.48	2.72	-2	-7	8	0.001
Patient 0-10 rating of mouth sensitivity	32	-1.28	3.59	-2	-7	10	0.004
Patient 0-10 rating of eye problems	38	**0.08**	3.80	0	-6	9	0.93

*Cases unaffected at both baseline and 1 year are excluded.

† Worsened measures are highlighted in bold (not statistically significant).

‡ signed rank test

List of Appendices

I. 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. http://www.bloodjournal.org/content/57/2/267

II. Wolff D, Gerbitz A, Ayuk F, et al. Consensus Conference on Clinical Practice in Chronic Graft-versus-Host Disease (GVHD): First-Line and Topical Treatment of Chronic GVHD. Biol Blood Marrow Transplant 2010; 16(12):1611-1628. http://dx.doi.org/10.1016/j.bbmt.2010.06.015

III. Wolff D, Schleuning M, von Harsdorf S, et al. Consensus Conference on Clinical Practice in Chronic GVHD: Second-Line Treatment of Chronic Graft-versus-Host Disease. Biol Blood Marrow Transplant 2011;17(1):1-17. http://www.sciencedirect.com/science/article/pii/S1083879110002235

IV. Vigorito AC, Campregher PV, Storer BE, et al. Evaluation of NIH consensus criteria for classification of late acute and chronic GVHD. *Blood*. 2009;114(3):702-708. http://www.bloodjournal.org/content/114/3/702

V. Martin PJ, Weisdorf D, Przepiorka D, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: VI. Design of Clinical Trials Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(5):491-505. http://www.sciencedirect.com/science/article/pii/S1083879106002503

VI. Inamoto Y, Flowers MED, Sandmaier BM, et al. Failure-free survival after initial systemic treatment of chronic graft-versus-host disease. *Blood.* 2014;124(8):1363-1371. http://www.bloodjournal.org/content/124/8/1363

VII. Consortium CG. Rationale and design of the chronic GVHD cohort study: improving outcomes assessment in chronic GVHD. *Biol Blood Marrow Transplant*. 2011;17(8):1114-1120. http://www.sciencedirect.com/science/article/pii/S108387911100214X

VIII. Provider Instrument used in prospective, longitudinal study (attached)

IX. Patient instrument used in prospective, longitudinal study (attached)

X. Lee S, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2002;8(8):444-452. http://www.sciencedirect.com/science/article/pii/S1083879102500185

XI. Lee SJ, Wolff D, Kitko D, et al. Measuring therapeutic response in chronic graft-versus-host disease. National Institutes of Health Consensus Development Project on Criteria for Clinical

Trials in Chronic Graft-Versus-Host Disease: IV. The 2014 Response Criteria Working Group Report. *Biol Blood Marrow Transplant*. 2015; in press. http://www.sciencedirect.com/science/article/pii/S108387911500155X

XII. Inamoto Y, Storer BE, Lee SJ, et al. Failure-free survival after second-line systemic treatment of chronic graft-versus-host disease. *Blood*. 2013;121(12):2340-2346. http://www.bloodjournal.org/content/121/12/2340

XIII. Jacobsohn DA. FFS: an end(point) to our problems in chronic GVHD trials? Blood 2014:124(8):1216-1217. http://www.bloodjournal.org/content/124/8/1216

XIV. Vogelsang G, Order out of chaos. Blood 2013;121(12):2170-2172. http://www.bloodjournal.org/content/121/12/2170

XV. Examples of potential development paths (attached)

Chronic GVHD Protocol

Improving outcomes assessment in chronic GVHD

Provider Survey

Follow-up

Instructions:
Please score a symptom only if you know or suspect it be *related to chronic GVHD*.
Subjective symptoms are acceptable. For example, joint tightness can be scored based on subjective findings despite the absence of objective limitations.

Please score symptoms present in the *last week*. Even if they may have resolved with treatment in the past week, if they were present recently and may possibly return, please score them.

Date of Visit: _____

Patient: _____

MRN: _____

Your Name: _____

Target Date for Next Visit (+/- 1 mo): _____

Confirmed Next Visit Date: _____

SKIN

Do not use Rule of 9s *Indicate % of body part affected*		Erythematous rash of any sort	Moveable sclerosis	Non-moveable subcutaneous sclerosis or fasciitis
1. Head/neck/scalp		%	%	%
2. Anterior torso		%	%	%
3. Posterior torso		%	%	%
4. L. upper extremity		%	%	%
5. R. upper extremity		%	%	%
6. L. lower extremity, (incl. L buttock)		%	%	%
7. R. lower extremity, (incl. R buttock)		%	%	%
8. Genitalia	❑ not examined	%	%	%

	0	1	2	3	4
Skin sclerotic changes	❑ Normal	❑ Thickened with pockets of normal skin	❑ Thickened over majority of skin	❑ Thickened, unable to move	❑ Hidebound, unable to pinch

	0	**1**	**2**	**3**
Skin Score	❑ No Symptoms	❑ <18% BSA with disease signs but **NO** sclerotic features	❑ 19-50% BSA **OR** involvement with superficial sclerotic features "not hidebound" (able to pinch)	❑ >50% BSA **OR** deep sclerotic features "hidebound" (unable to pinch) **OR** impaired mobility, ulceration or severe pruritus
Fascia	❑ Normal	❑ Tight with normal areas	❑ Tight	❑ Tight, unable to move

Clinical Skin Features

❑ Ulcer Location: _____ Largest dimension: _____ cm

❑ Maculopapular rash	❑ Keratosis pilaris
❑ Lichen planus-like lesions	❑ Papulosquamous lesions or icthyosis
❑ Poikiloderma	❑ Hair involvement
❑ Pruritus	❑ Nail involvement
❑ Other, specify:	❑ Other, specify:

SKIN

Region	Grade	% Area of Grade	Fraction of Grade 3 or 4 Areas with Erythema (indicate up to what fraction is involved)	Region	Grade	% Area of Grade	Fraction of Grade 3 or 4 Areas with Erythema (indicate up to what fraction is involved)
1. Head, Neck and Scalp	0	%		6. Right Hand	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	**100 %**			**Total =**	**100 %**	
2. Chest	0	%		7. Left Arm	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	**100 %**			**Total =**	**100 %**	
3. Abdomen and Genitals	0	%		8. Left Hand	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	**100 %**			**Total =**	**100 %**	
4. Back and Buttocks	0	%		9. Right Leg and Foot	0	%	
	1	%			1	%	
	2	%			2	%	
	3	%	❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	**100 %**			**Total =**	**100 %**	
5. Right Arm	0			10. Left Leg and Foot	0		
	1				1		
	2				2		
	3		❑0 ❑¼ ❑½ ❑¾ ❑1		3	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	4	%	❑0 ❑¼ ❑½ ❑¾ ❑1		4	%	❑0 ❑¼ ❑½ ❑¾ ❑1
	Total =	**100 %**			**Total =**	**100 %**	

Percentages must add up to 100

0 = normal skin

1 = discolored [hypopigmentation, hyperpigmentation, alopecia, erythema, maculopapular rash]

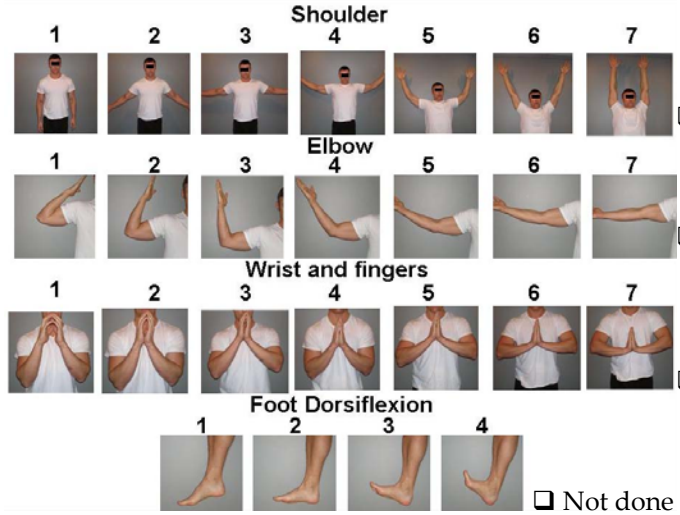
2 = lichenoid plaque, or skin thickened (able to move)

3 = skin thickened with limited motion but able to pinch [scleroderma or fasciae involvement]

4 = hidebound skin, unable to move, unable to pinch

ROM & MOUTH

Please circle this person's current ROM for each joint from 1=poor mobility to 7=full mobility below:



❑ Not done

❑ Not done

❑ Not done

❑ Not done

		0	1	2	3
Mouth Score		❑ No symptoms	❑ Mild symptoms with disease signs but not limiting oral intake significantly	❑ Moderate symptoms with signs with **partial** limitation of oral intake	❑ Severe symptoms with disease signs on examination with **major** limitation of oral intake
Mouth	**Erythema**	❑ None	❑ Mild erythema OR Moderate erythema (<25%)	❑ Moderate (≥25%) OR Severe erythema (<25%)	❑ Severe erythema (≥25%)
	Lichenoid	❑ None	❑ Hyperkeratotic changes (<25%)	❑ Hyperkeratotic changes (25-50%)	❑ Hyperkeratotic changes (>50%)
	Ulcers	❑ None	❑ None	❑ Ulcers involving (≤20%)	❑ Severe ulcerations (>20%)
	Mucoceles (of lower labia and soft palate only)	❑ None	❑ 1-5 mucoceles	❑ 6-10 scattered mucoceles	❑ Over 10 mucoceles
Mouth Pain		❑ No symptoms	❑ Food sensitivity	❑ Pain requiring narcotics	❑ Unable to eat

GASTROINTESTINAL

		0	1	2	3
GI Tract Score		❑ No symptoms	❑ Symptoms such as dysphagia, anorexia, nausea, vomiting, abdominal pain or diarrhea without significant weight loss (<5%)	❑ Symptoms associated with mild to moderate weight loss (5-15%)	❑ Symptoms associated with significant weight loss >15%, requires nutritional supplement for most calorie needs **OR** esophageal dilation
Gastro-intestinal	**Esophagus** • **Dysphagia OR** • **Odynophagia**	❑ No esophageal symptoms	❑ Occasional dysphagia or odynophagia with solid food or pills *during the past week*	❑ Intermittent dysphagia or odynophagia with solid food or pills (but not for liquids or soft foods) *during the past week*	❑ Dysphagia or odynophagia for almost all oral intake, *on almost every day of the past week*
	Upper GI • **Early satiety OR** • **Anorexia OR** • **Nausea & vomiting**	❑ No symptoms	❑ Mild, occasional symptoms with little reduction in oral intake *during the past week*	❑ Moderate, intermittent symptoms throughout the day, with some reduction in oral intake, *during the past week*	❑ More severe or persistent symptoms throughout the day, with marked reduction in oral intake, *on almost every day of the past week*
	Lower GI • **Diarrhea**	❑ No loose or liquid stools *during the past week*	❑ Occasional loose or liquid stools, on some days *during the past week*	❑ Intermittent loose or liquid stools through-out the day, *on almost every day of the past week* **without requiring** intervention to prevent or correct volume depletion	❑ Voluminous diarrhea *on almost every day of the past week* **requiring** intervention to prevent or correct volume depletion

OTHER ORGANS

	0	1	2	3
Eye Score	❑ No symptoms	❑ Mild dry eye symptoms not affecting ADL (requiring eye drops <3x per day) **OR** asymptomatic signs of kerato-conjunctivitis sicca	❑ Moderate dry eye symptoms partially affecting ADL (requiring eye drops >3x per day or punctual plugs) **WITHOUT** vision impairment	❑ Severe dry eye symptoms significantly affecting ADL (special eyewear to relieve pain) **OR** unable to work because of ocular symptoms **OR** loss of vision caused by kerato-conjunctivitis sicca
Joints and Fascia Score	❑ No symptoms	❑ Mild tightness of arms or legs, normal or mild decreased range of motion (ROM) **AND** not affecting ADL	❑ Tightness of arms or legs **OR** joint contractures, erythema thought due to fasciitis, moderate decrease ROM **AND** mild to moderate limitation of ADL	❑ Contracture WITH significant decrease of ROM **AND** significant limitation of ADL (unable to tie shoes, button shirts, dress self etc.)
Genital Tract Score **(score even if no GYN exam, required for men too)** ❑ **No GYN Exam**	❑ No symptoms	❑ Symptomatic with mild distinct signs on exam AND no effect on coitus and minimal discomfort with GYN exam	❑ Symptomatic with distinct signs on exam AND with mild dyspareunia or discomfort with GYN exam	❑ Symptomatic WITH advanced signs (stricture, labia agglutination or severe ulceration) AND severe pain with coitus or inability to insert vaginal spectrum
Lung Score	❑ No symptoms	❑ Mild symptoms (shortness of breath after climbing one flight of steps)	❑ Moderate symptoms (shortness of breath after walking on flat ground)	❑ Severe symptoms (shortness of breath at rest; requiring O_2)
Other Organ Score *Specify:* _____	❑ No effect on ADL	❑ Mild effect on ADL	❑ Moderate effect on ADL	❑ Severe effect on ADL
Other Organ Score *Specify:* _____	❑ No effect on ADL	❑ Mild effect on ADL	❑ Moderate effect on ADL	❑ Severe effect on ADL

OVERALL STATUS

Please rate the severity of this person's chronic GVHD	
on this scale →	❑ None (0) ❑ Mild (1) ❑ Moderate (2) ❑ Severe (3)
and on this scale → *(circle one)*	cGVHD symptoms are not at all severe ⟷ cGVHD symptoms are most severe possible 0 1 2 3 4 5 6 7 8 9 10

Current GVHD Status	❑ Complete response (1)	❑ Partial response (2)	❑ Unchanged (3)	❑ Progressive (4)

Reasons for changing therapeutic regimen (Check all that apply)	
❑ Not applicable, no changes made	❑ Toxicity
❑ Adjust levels of medications	❑ New symptoms
❑ Enroll on clinical trial	❑ Improvement in symptoms
❑ Worsening of symptoms	❑ Disease relapse
❑ No improvement in symptoms	❑ Stable

Does this person *currently* have:	❑ Late acute GVHD (1) ❑ Overlap acute and chronic GVHD (2) ❑ Classic chronic GVHD (3) ❑ No GVHD (0)

OVERALL STATUS

Since the last study visit six months ago on _____, how would you say this patient's chronic GVHD has changed?

	Not involved (0)	Resolved (1)	Very much better (2)	Moderately better (3)	A little better (4)	About the same (5)	A little worse (6)	Moderately worse (7)	Very much worse (8)
Mouth	❑	❑	❑	❑	❑	❑	❑	❑	❑
Skin	❑	❑	❑	❑	❑	❑	❑	❑	❑
Eyes	❑	❑	❑	❑	❑	❑	❑	❑	❑
Joints	❑	❑	❑	❑	❑	❑	❑	❑	❑
Chronic GVHD Overall		❑	❑	❑	❑	❑	❑	❑	❑

What are your reasons for how you rated "chronic GVHD overall"? *Write in* ➔ (For example, has an organ or symptom improved or worsened?)	

Infection	**0**	**1**	**2**	**3**	**4**
	❑ None	❑ Mild, topical or no therapy required	❑ Moderate, localized, requiring oral treatment For 2-4:	❑ Severe, systemic infection requiring IV anti-infective, mold-active oral antifungal or hospitalization	❑ Life-threatening infection
		❑ Pending lab report (1)	❑ Unidentified organism (2) ❑ Identified organism, specify (3):		

Peripheral Edema?	❑ None (0)	❑ Tr (9)	❑ 1+	❑ 2+	❑ 3+	❑ 4+

OVERALL STATUS

Other indicators, clinical manifestations or severe complications related to chronic GVHD					
	Never (0)	**Past, not now (1)**	**Mild (2)**	**Moderate (3)**	**Severe (4)**
1. Pleural Effusion(s)	❑	❑	❑	❑	❑
2. Bronchiolitis obliterans	❑	❑	❑	❑	❑
3. Bronchiolitis obliterans organizing pneumonia	❑	❑	❑	❑	❑
4. Nephrotic syndrome	❑	❑	❑	❑	❑
5. Malabsorption	❑	❑	❑	❑	❑
6. Esophageal stricture or web	❑	❑	❑	❑	❑
7. Ascites (serositis)	❑	❑	❑	❑	❑
8. Myasthenia Gravis	❑	❑	❑	❑	❑
9. Peripheral Neuropathy	❑	❑	❑	❑	❑
10. Polymyositis	❑	❑	❑	❑	❑
11. Pericardial Effusion	❑	❑	❑	❑	❑
12. Cardiomyopathy	❑	❑	❑	❑	❑
13. Cardiac conduction defects	❑	❑	❑	❑	❑
14. Coronary artery involvement	❑	❑	❑	❑	❑
15. Other, please specify: _____	❑	❑	❑	❑	❑
16. Other, please specify: _____	❑	❑	❑	❑	❑
17. Other, please specify: _____	❑	❑	❑	❑	❑

For office use only:

Study ID	*Initials (First, Last)*	*Date completed:*	*Date received:*
Person completing form:		*Their degree:*	
Timepoint: *v3.3*		*Date entered:*	

Chronic GVHD Patient Survey

FOLLOW-UP

INSTRUCTIONS

This survey will provide us with important information about your health.

All your answers will be kept strictly confidential and will not be included in your medical record. The information that you provide will be combined with that of many other transplant patients before analysis.

Please read each question carefully. Circle or check off the answer that best describes how you feel.

While we ask that you answer each question, you are free to *not* answer any question that makes you feel uncomfortable. If none of the answers provided seems exactly right, choose the one that comes closest to being right for you. Some of the questions may seem the same. However, it is important that we ask about certain aspects of your health in different ways in order to fully understand how you are feeling.

When you have completed this survey, please give it back to the study coordinator or mail it back to us using the enclosed self-addressed, stamped envelope.

We greatly appreciate your participation.

Your name: _____ Date: _____

Section 1: Your Chronic Graft vs. Host Disease (GVHD) Symptoms

Please circle the number that shows how severe your symptoms have been **in the last week**:

Not Present ⟵ **As Bad As You Can Imagine**

1.	Your chronic GVHD symptoms **overall**?	0	1	2	3	4	5	6	7	8	9	10
2.	Your **skin itching** at its WORST?	0	1	2	3	4	5	6	7	8	9	10
3.	Your **mouth dryness** at its WORST?	0	1	2	3	4	5	6	7	8	9	10
4.	Your **mouth pain** at its WORST?	0	1	2	3	4	5	6	7	8	9	10
5.	Your **mouth sensitivity** at its WORST?	0	1	2	3	4	5	6	7	8	9	10
6.	Your **eye** problem at its WORST?	0	1	2	3	4	5	6	7	8	9	10

7. What is your main complaint with regard to your eyes?

   (Write in): _____

8. ***Vulvovaginal Symptoms (females only):*** Do you have any burning, pain or discomfort in the area of your vagina, vulva or labia? - **OR -**
   Do you have any discomfort or pain with sexual intercourse?

   1-☐ Yes 0-☐ No -2-☐ Not applicable

9. ***(Male and female)*** Overall, how would you rate the severity of your chronic graft versus host disease?

   0-☐ None
   1-☐ Mild
   2-☐ Moderate
   3-☐ Severe

10. Do you think your chronic GVHD symptoms are in good enough control to decrease your immunosuppressive medications?

    0-☐ No
    1-☐ Yes
    -2-☐ Not applicable

Compared to your last study visit six months ago on _____, how would you rate your following GVHD symptoms now?

		Not involved with GVHD	Compl-etely gone	Very much better	Moder-ately better	A little better	About the same	A little worse	Moder-ately worse	Very much worse
11.	**GVHD symptoms overall**	--	1	2	3	4	5	6	7	8
12.	**Mouth**	0	1	2	3	4	5	6	7	8
13.	**Skin**	0	1	2	3	4	5	6	7	8
14.	**Eye**	0	1	2	3	4	5	6	7	8
15.	**Joints**	0	1	2	3	4	5	6	7	8

16. What are your reasons for saying your chronic GVHD is better or worse overall? (Is there a symptom of particular concern to you that has changed?)

(Write in)

_____

_____

_____

_____

_____

_____

Please continue to next page ☞

Section 2

By circling one (1) number per line, please indicate how much you have been bothered by the following problems in the past month:

SKIN:		**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
1.	Abnormal skin color……..	0	1	2	3	4
2.	Rashes……	0	1	2	3	4
3.	Thickened skin……. ……...	0	1	2	3	4
4.	Sores on skin……	0	1	2	3	4
5.	Itchy skin…… …….	0	1	2	3	4

EYES AND MOUTH:		**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
6.	Dry eyes……	0	1	2	3	4
7.	Need to use eye drops frequently..	0	1	2	3	4
8.	Difficulty seeing clearly……	0	1	2	3	4
9.	Need to avoid certain foods due to mouth pain……	0	1	2	3	4
10.	Ulcers in mouth……	0	1	2	3	4
11.	Receiving nutrition from an intravenous line or feeding tube....	0	1	2	3	4

BREATHING:		**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
12.	Frequent cough……	0	1	2	3	4
13.	Colored sputum……	0	1	2	3	4
14.	Shortness of breath with exercise..	0	1	2	3	4
15.	Shortness of breath at rest……..	0	1	2	3	4
16.	Need to use oxygen……..	0	1	2	3	4

EATING AND DIGESTION:	**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
17. Difficulty swallowing solid foods….	0	1	2	3	4
18. Difficulty swallowing liquids…….	0	1	2	3	4
19. Vomiting……	0	1	2	3	4
20. Weight loss……	0	1	2	3	4

MUSCLES AND JOINTS:	**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
21. Joint and muscle aches……..	0	1	2	3	4
22. Limited joint movement……	0	1	2	3	4
23. Muscle cramps……..	0	1	2	3	4
24. Weak muscles……..	0	1	2	3	4

ENERGY:	**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
25. Loss of energy……	0	1	2	3	4
26. Need to sleep more/take naps…..	0	1	2	3	4
27. Fevers……	0	1	2	3	4

MENTAL AND EMOTIONAL:	**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
28. Depression……	0	1	2	3	4
29. Anxiety……..	0	1	2	3	4
30. Difficulty sleeping……	0	1	2	3	4

Section 3

Have you experienced any of the following during the last week?		All of the time	Most of the time	Half of the time	Some of the time	None of the time
1.	Eyes that are sensitive to light?.....	4	3	2	1	0
2.	Eyes that feel gritty?.....	4	3	2	1	0
3.	Painful or sore eyes?.....	4	3	2	1	0
4.	Blurred vision?.....	4	3	2	1	0
5.	Poor vision?.....	4	3	2	1	0

Have problems with your eyes limited you in performing any of the following during the last week?		All of the time	Most of the time	Half of the time	Some of the time	None of the time	Not applicable
6.	Reading? ……..	4	3	2	1	0	N/A (-2)_
7.	Driving at night? ……..	4	3	2	1	0	N/A (-2)_
8.	Working with a computer or bank machine (ATM)? …….	4	3	2	1	0	N/A (-2)_
9.	Watching TV? ……..	4	3	2	1	0	N/A (-2)_

Have your eyes felt uncomfortable in any of the following situations during the last week?		All of the time	Most of the time	Half of the time	Some of the time	None of the time	Not applicable
10.	Windy conditions?.....	4	3	2	1	0	N/A (-2)_
11.	Places or areas with low humidity (very dry)?.....	4	3	2	1	0	N/A (-2)_
12.	Areas that are air conditioned?.....	4	3	2	1	0	N/A (-2)_

Section 4: Quality of Your Life After Your Transplant

By circling one (1) number per line, please indicate how true each statement has been for you during the past 7 days:

PHYSICAL WELL-BEING:		Not at all	A little bit	Some-what	Quite a bit	Very much
1.	I have a lack of energy	0	1	2	3	4
2.	I have nausea	0	1	2	3	4
3.	Because of my physical condition, I have trouble meeting the needs of my family	0	1	2	3	4
4.	I have pain	0	1	2	3	4
5.	I am bothered by side effects of treatment	0	1	2	3	4
6.	I feel ill	0	1	2	3	4
7.	I am forced to spend time in bed	0	1	2	3	4

By circling one (1) number per line, please indicate how true each statement has been for you during the past 7 days:

SOCIAL/FAMILY WELL-BEING:	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
8. I feel close to my friends	0	1	2	3	4
9. I get emotional support from my family	0	1	2	3	4
10. I get support from my friends	0	1	2	3	4
11. My family has accepted my illness.....	0	1	2	3	4
12. I am satisfied with family communication about my illness	0	1	2	3	4
13. I feel close to my partner (or the person who is my main support)	0	1	2	3	4

Regardless of your current level of sexual activity, please answer the following question. If you prefer not to answer it, please check this box and go to the next section. ☐ **-3**

14. I am satisfied with my sex life ……	0	1	2	3	4

EMOTIONAL WELL-BEING:	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
15. I feel sad.....	0	1	2	3	4
16. I am satisfied with how I am coping with my illness	0	1	2	3	4
17. I am losing hope in the fight against my illness	0	1	2	3	4
18. I feel nervous.....	0	1	2	3	4
19. I worry about dying.....	0	1	2	3	4
20. I worry that my condition will get worse.....	0	1	2	3	4

By circling one (1) number per line, please indicate how true each statement has been for you during the past 7days.

FUNCTIONAL WELL-BEING:		**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
21.	I am able to work (include work at home)	0	1	2	3	4
22.	My work (include work at home) is fulfilling.....	0	1	2	3	4
23.	I am able to enjoy life	0	1	2	3	4
24.	I have accepted my illness	0	1	2	3	4
25.	I am sleeping well.....	0	1	2	3	4
26.	I am enjoying the things I usually do for fun.....	0	1	2	3	4
27.	I am content with the quality of my life right now.....	0	1	2	3	4

ADDITIONAL CONCERNS:		**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
28.	I am concerned about keeping my job (include work at home).....	0	1	2	3	4
29.	I feel distant from other people.....	0	1	2	3	4
30.	I worry that the transplant will not work	0	1	2	3	4
31.	The effects of treatment are worse than I had imagined.....	0	1	2	3	4
32.	I have a good appetite.....	0	1	2	3	4
33.	I like the appearance of my body	0	1	2	3	4
34.	I am able to get around by myself	0	1	2	3	4
35.	I get tired easily	0	1	2	3	4
36.	I am interested in sex	0	1	2	3	4
37.	I have confidence in my nurse(s)	0	1	2	3	4

Section 5: Your Health and Well-Being

This section asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities. Please answer every question. If you are unsure about how to answer a question, please give the answer that seems closest to how you feel.

For each of the following questions, please mark an ☒ in the one box that best describes your answer.

1. Which statement describes how you feel most of the time? (please check one)

- ☐ 1. Normal, no difficulties with daily activities
- ☐ 2. Able to carry on normal activities, minor problems
- ☐ 3. Normal activity with effort
- ☐ 4. Able to care for self, but unable to carry on normal activity or active work
- ☐ 5. Require occasional assistance, but able to care for most of needs
- ☐ 6. Require considerable assistance and frequent medical care
- ☐ 7. Disabled, require special care and assistance
- ☐ 8. Severely disabled, hospitalized
- ☐ 9. Very sick, hospitalized

2. In general, would you say your health is:

Excellent	Very good	Good	Fair	Poor
▼	▼	▼	▼	▼
☐1	☐2	☐3	☐4	☐5

3. Compared to one year ago, how would you rate your health in general now?

Much better now than one year ago	Somewhat better now than one year ago	About the same as one year ago	Somewhat worse now than one year ago	Much worse now than one year ago
▼	▼	▼	▼	▼
☐1	☐2	☐3	☐4	☐5

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes, limited a lot	Yes, limited a little	No, not limited at all
4. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
5. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
6. Lifting or carrying groceries	1	2	3
7. Climbing several flights of stairs	1	2	3
8. Climbing one flight of stairs	1	2	3
9. Bending, kneeling, or stooping	1	2	3
10. Walking more than a mile	1	2	3
11. Walking several hundred yards	1	2	3
12. Walking one hundred yards	1	2	3
13. Bathing or dressing yourself	1	2	3

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
14. Cut down on the amount of time you spent on work or other activities	1	2	3	4	5
15. Accomplished less than you would like	1	2	3	4	5
16. Were limited in the kind of work or other activities	1	2	3	4	5
17. Had difficulty performing the work or other activities (for example, it took extra effort)	1	2	3	4	5

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
18. Cut down on the amount of time you spent on work or other activities	1	2	3	4	5
19. Accomplished less than you would like	1	2	3	4	5
20. Did work or other activities less carefully than usual	1	2	3	4	5

21. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐1	☐2	☐3	☐4	☐5

22. How much bodily pain have you had during the past 4 weeks?

None	Very mild	Mild	Moderate	Severe	Very Severe
☐1	☐2	☐3	☐4	☐5	☐6

23. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐1	☐2	☐3	☐4	☐5

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks...

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
24. Did you feel full of life?	1	2	3	4	5
25. Have you been very nervous?	1	2	3	4	5
26. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5
27. Have you felt calm and peaceful?	1	2	3	4	5
28. Did you have a lot of energy?	1	2	3	4	5
29. Have you felt downhearted and depressed?	1	2	3	4	5
30. Did you feel worn out?	1	2	3	4	5
31. Have you been happy?	1	2	3	4	5
32. Did you feel tired?	1	2	3	4	5

33. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
▼	▼	▼	▼	▼
☐1	☐2	☐3	☐4	☐5

How TRUE or FALSE is each of the following statements for you?

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false
	▼	▼	▼	▼	▼
34. I seem to get sick a little easier than other people	1	2	3	4	5
35. I am as healthy as anybody I know	1	2	3	4	5
36. I expect my health to get worse	1	2	3	4	5
37. My health is excellent	1	2	3	4	5

Section 6: Your Activity Level

Please check each activity according to these directions:

Check Column 1 ("Still Doing This Activity") if you completed the activity unassisted the last time you had the need or opportunity to do so.

Check Column 2 ("Have Stopped Doing This Activity") if you have engaged in the activity in the past, but you probably would not perform the activity today even if the opportunity should arise.

Check Column 3 ("Never Did This Activity") if you have never engaged in the specific activity.

	Still Doing This Activity (1)	Have Stopped Doing This Activity (2)	Never Did This Activity (3)
1. Getting in and out of chairs or bed (without assistance)			
2. Listening to the radio			
3. Reading books, magazines or newspapers			
4. Writing (letters, notes)			
5. Working at a desk or table			
6. Standing (for more than one minute)			
7. Standing (for more than five minutes)			
8. Dressing or undressing (without assistance)			
9. Getting clothes from drawers or closets			
10. Getting in or out of a car (without assistance)			
11. Dining at a restaurant			
12. Playing cards/table games			
13. Taking a bath (no assistance needed)			
14. Putting on shoes, stockings or socks (no assistance needed)			
15. Attending a movie, play, church event or sports activity			
16. Walking 30 yards (27 meters)			

	Still Doing This Activity (1)	Have Stopped Doing This Activity (2)	Never Did This Activity (3)
17. Walking 30 yards (non-stop)			
18. Dressing/undressing (no rest or break needed)			
19. Using public transportation or driving a car (100 miles or less)			
20. Using public transportation or driving a car (99 miles or more)			
21. Cooking your own meals			
22. Washing or drying dishes			
23. Putting groceries on shelves			
24. Ironing or folding clothes			
25. Dusting/polishing furniture or polishing cars			
26. Showering			
27. Climbing six steps			
28. Climbing six steps (non-stop)			
29. Climbing nine steps			
30. Climbing 12 steps			
31. Walking ½ block on level ground			
32. Walking ½ block on level ground (non-stop)			
33. Making a bed (not changing sheets)			
34. Cleaning windows			
35. Kneeling, squatting to do light work			
36. Carrying a light load of groceries			
37. Climbing nine steps (non-stop)			

	Still Doing This Activity (1)	Have Stopped Doing This Activity (2)	Never Did This Activity (3)
38. Climbing 12 steps (non-stop)			
39. Walking ½ block uphill			
40. Walking ½ block uphill (non-stop)			
41. Shopping (by yourself)			
42. Washing clothes (by yourself)			
43. Walking one block on level ground			
44. Walking two blocks on level ground			
45. Walking one block on level ground (non-stop)			
46. Walking two blocks on level ground (non-stop)			
47. Scrubbing (floors, walls or cars)			
48. Making beds (changing sheets)			
49. Sweeping			
50. Sweeping (five minutes non-stop)			
51. Carrying a large suitcase or bowling (one line)			
52. Vacuuming carpets			
53. Vacuuming carpets (five minutes non-stop)			
54. Painting (interior/exterior)			
55. Walking six blocks on level ground			
56. Walking six blocks on level ground (non-stop)			
57. Carrying out the garbage			
58. Carrying a heavy load of groceries			

	Still Doing This Activity (1)	Have Stopped Doing This Activity (2)	Never Did This Activity (3)
59. Climbing 24 steps			
60. Climbing 36 steps			
61. Climbing 24 steps (non-stop)			
62. Climbing 36 steps (non-stop)			
63. Walking one mile			
64. Walking one mile (non-stop)			
65. Running 110 yards (100 meters) or playing softball/baseball			
66. Dancing (social)			
67. Doing calisthenics or aerobic dancing (5 minutes non-stop)			
68. Mowing the lawn (power mower, but not a riding mower)			
69. Walking two miles			
70. Walking two miles (non-stop)			
71. Climbing 50 steps			
72. Shoveling, digging or spading			
73. Shoveling, digging or spading (five minutes non-stop)			
74. Climbing 50 steps (non-stop)			
75. Walking three miles or golfing 18 holes without a riding cart			
76. Walking three miles (non-stop)			
77. Swimming 25 yards			
78. Swimming 25 yards (non-stop)			

	Still Doing This Activity (1)	Have Stopped Doing This Activity (2)	Never Did This Activity (3)
79. Bicycling one mile			
80. Bicycling two miles			
81. Bicycling one mile (non-stop)			
82. Bicycling two miles (non-stop)			
83. Running or jogging ¼ mile			
84. Running or jogging ½ mile			
85. Playing tennis or racquetball			
86. Playing basketball (game play)			
87. Running or jogging ¼ mile (non-stop)			
88. Running or jogging ½ mile (non-stop)			
89. Running or jogging one mile			
90. Running or jogging two miles			
91. Running or jogging three miles			
92. Running or jogging one mile in 12 minutes or less			
93. Running or jogging two miles in 20 minutes or less			
94. Running or jogging three miles in 30 minutes or less			

Section 7: About Yourself

1. What is your current work status? (circle **all** that apply)

In school full time.....1
In school part time.....2
Working full time.....3
Working part time.....4
Homemaker.....5
Retired.....6
On medical leave from work.....7
Disabled, unable to work.....8
Unemployed, looking for work.....9
Unemployed, not looking for work... 10
Other, specify..... 12

Thank you for participating in this study

Please remember that someone is available to speak with you at any time, if you wish. Dr. <<Site PI>> may be reached by calling <<Site PI phone>>. (S)he will be able to answer any questions about the study or refer you to other support staff as needed.

Please use the space below for any other comments.

For office use only:

Study ID	*Initials (First, Last)*	*Date completed:*	*Date received:*
Timepoint: *V3.0*		*Date entered:*	

Examples of Possible Development Paths for Investigational Products Intended for Treatment of Chronic GVHD

Development paths leading to regulatory approval for indications related to chronic GVHD have not been established. The small market and lack of an established development path stand as disincentives for industry sponsors. Establishment of development paths could decrease the risks for industry sponsors and increase their interest in chronic GVHD. As examples, we outline development paths for products intended to provide treatment effects for 3 indications related to chronic GVHD: 1) systemic control of disease manifestations that require immediate intervention as second-line treatment, 2) systemic control of mild to moderate manifestations at initial diagnosis, and 3) local control of specific disease manifestations (Figure 1). These examples are not intended to be comprehensive, and they do not address all possible contingencies.

Abbreviations: FFS, failure-free survival; SWOPI, survival without progressive impairment

Systemic control of disease manifestations that require immediate intervention as second-line treatment. The first trial in this pathway could be designed to test the efficacy of a product for controlling rapidly reversible manifestations of disease activity in patients with chronic GVHD that has not responded adequately to initial treatment, including glucocorticoids. Manifestations of disease activity include erythematous rash, oral mucosal changes, conjunctival inflammation not caused by dry eye, abnormal liver function tests, and rapidly reversible gastrointestinal manifestations, including nausea, vomiting, diarrhea and weight loss. Baseline treatment with agents other than glucocorticoids may be continued, but no new systemic or topically active agents should be added, other than the investigational product. Improvement in these manifestations of disease activity should generally be evident within 4 weeks, but longer durations of administration will be needed even in the initial studies in order to assess the durability of response, the effects on less rapidly reversible manifestations such as weight loss, sclerosis and fasciitis, the ability to taper concomitant medications, including the dose of glucocorticoids, and the safety of the product in patients with chronic GVHD. Longer durations of administration may also be needed before the response assessment when products are known not to produce prompt improvement (e.g., extra-corporeal photopheresis).

If the first trial shows evidence of efficacy in controlling manifestations of disease activity, a follow-up or extension phase II trial could address the question of whether the product improves FFS or SWOPI in patients with chronic GVHD that has not responded adequately to initial treatment. In all trials using FFS or SWOPI as the primary endpoint, response outcomes should be defined, measured, documented and reported, and serial steroid dose data should be collected. Successful results could lead to a controlled phase III trial for the same indication.

If the follow-up or extension trial improves FFS or SWOPI, the development path could shift to initial treatment. It is also possible that the development path could bypass the first phase of testing with second-line treatment and begin with initial treatment (see below).

While a major purpose of phase II studies is to identify promising approaches for phase III studies, another important purpose of phase II studies is to identify approaches that do not work, in order to

avoid unnecessary investments in large expensive phase III studies that have little chance of success. FFS, SWOPI or durable response could be very effectively used in multi-stage trial designs. The first stage would enroll patients in a traditional single-arm phase II study with FFS, SWOPI or durable response at 6 months or perhaps a year as the primary endpoint compared against historical results for the same indication, with important secondary endpoints such as survival, relapse and non-relapse mortality. A sample size of 40 to 60 patients would suffice for this purpose. Positive results would trigger an immediate second stage, but the phase III stage would have to "start over" with enrollment, excluding results from the first stage in order to avoid bias. FFS, SWOPI, long-term durable response or clinical tolerance could be used as the primary endpoint. For this purpose, clinical tolerance is defined as resolution of GVHD without resorting to another treatment, followed by withdrawal of all systemic treatment and no subsequent recurrence of GVHD. Alternatively, a randomized phase III group-sequential design could be used with an FFS or SWOPI-based interim stopping rule for futility, using FFS, SWOPI or clinical tolerance as the primary endpoint.

Systemic control of mild or moderate manifestations at initial diagnosis. An investigational product could be adding to initial glucocorticoid treatment, using FFS, SWOPI or durable response as the primary endpoint in a single-arm phase II study. If results of a single-arm phase II study suggest success, a blinded randomized phase II study could be done to determine whether the postulated difference in FFS or SWOPI with and without the product added to initial glucocorticoid treatment could be confirmed. Such a study is likely to be underpowered, unless the difference between arms is very large. Nonetheless, results of this trial would be useful in providing true estimates of the difference between the arms and would provide longer-term safety data with the use of the product in patients with chronic GVHD.

Anti-inflammatory glucocorticoids have a long-established, prominent role in the treatment of chronic GVHD. Long-term, high-dose glucocorticoid treatment causes many side effects, some of which are irreversible. These considerations motivate interest in testing treatment regimens that do not contain anti-inflammatory glucocorticoids. Eligibility for such a trial would require the absence of current systemic glucocorticoid treatment. Initial systemic treatment for newly diagnosed chronic GVHD with mild manifestations represents the most likely setting in which the benefits and risks of a glucocorticoid-free regimen could be tested. Disease manifestations would have to be sufficiently severe to require systemic treatment, but not so severe as to require immediate use of glucocorticoids.

As a first step, an investigational product could be evaluated as a glucocorticoid-free treatment approach in a trial with FFS, SWOPI or durable response as the primary endpoint. Response should be measured at multiple time points in order to determine whether the benefit is sustained, and improvement should not be counted as a response if systemic or local treatment with glucocorticoids or another product has been added to the regimen before response is assessed. Given the uncertainty of benefit in such a trial, close monitoring at frequent intervals would be needed in order to ensure that symptoms and disease activity are adequately controlled and that no damage is emerging. Although no benchmark response rates have been established in this setting, a high rate of sustained responses would suggest that the product has activity against chronic GVHD, since previous results have indicated that untreated “clinical extensive” chronic GVHD tends to progress inexorably toward disability.

A pivotal randomized phase III study could be done if results of the first study with or without glucocorticoids suggest that the investigational treatment produced rates of response, FFS or SWOPI better than those expected with the standard of care. As discussed above, a multi-stage design could be used to improve the efficiency of the approach. A randomized phase III study could also be done if results of the initial study with a glucocorticoid-free regimen suggest that the investigational treatment produced rates of response, FFS or SWOPI equivalent to those expected for the standard of care. In this case, the primary endpoint would be designed to test the hypothesis that the overall burden of adverse effects is lower with the glucocorticoid-free investigational treatment than with the standard of care.

Local control of specific disease manifestations. Certain products might be suitable for treatment of specific systemic manifestations such as fibrosis, or might have local effect on specific organs such as the skin, mouth, eyes, lungs, gastrointestinal tract or genitourinary tract. Eligibility criteria for a trial to test a product for such an indication require careful definition to ensure that the condition is actually caused by chronic GVHD and that other potentially confounding causes are absent.

The first trial in the pathway could be designed to test the efficacy of a product for controlling rapidly reversible manifestations of disease activity, using an objective or subjective response measure as the primary endpoint. Claims that improvements after enrollment are related to the investigational treatment would be credible if the baseline systemic treatment is not changed at enrollment, and if no new systemic treatments are added before the assessment of response. In such a trial, the addition of a local or topical therapy before the assessment of response would count as failure. The first trial in the pathway could also be designed to test the efficacy of a product for preventing progression of less reversible manifestations of chronic GVHD. In this case, however, the trajectory of progression would have to be thoroughly documented before enrollment in order to determine whether a change occurred after enrollment in the study.

A pivotal randomized phase III study could be done if results of the first study suggest that the investigational treatment produced unambiguous sustained responses or prevented progression. In such a trial, one arm would be treated with the investigational product while continuing prior treatment without change, and the other arm would continue prior treatment without change. Such a trial should have a blinded design, if possible, in order to minimize bias. With blinding, a crossover design could be used to improve enrollment and motivate adherence to the protocol.

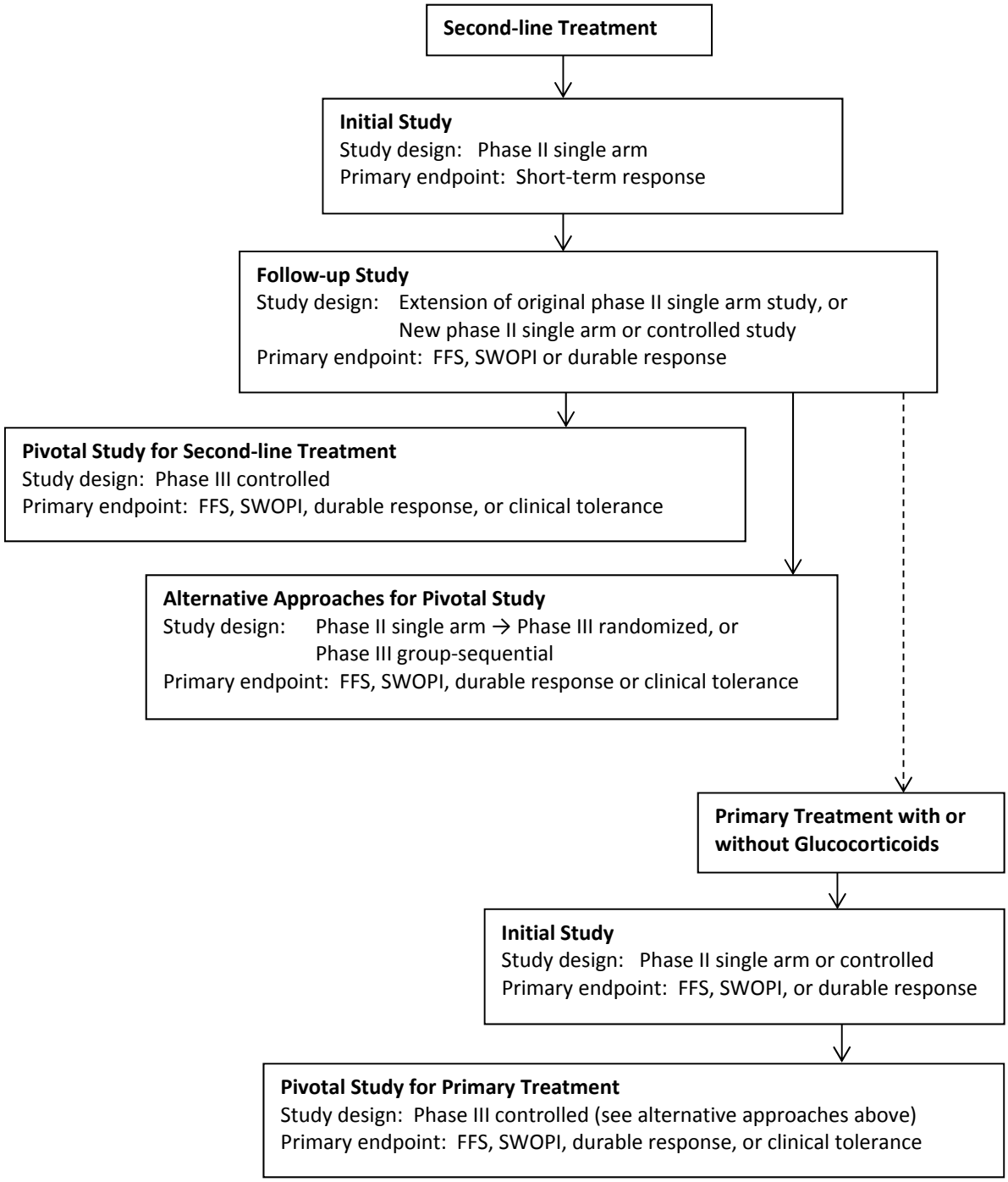


Figure 1. Overview of possible development paths for investigational products intended for systemic treatment of chronic GVHD.

Hello Dr. Martin,

In order for us to address your questions most completely, we would like to ask if you could please supply the following by Wednesday January 21, 2015:

1. The identity of the references for Supplementary Table 1 (the superscripts in the table do not match the reference list).

2. Overall survival (preferably KM curves) by NIH global severity score (mild, moderate, severe) from start of therapy for:
a) first line treatment with steroids alone
b) first line treatment with steroids plus a CNI
c) first salvage for patients refractory to steroids alone
d) first salvage for patients refractory to steroids plus a CNI

We understand that these may be very different patient groups, so feel free to describe briefly any difference in demographics as well.

3. Median (range) duration of treatment with immunosuppressive therapy (or time to discontinuation by CIF with death as a competing risk) for the same groups (a-d) listed in #2.

4. Are there any data available to determine the percentage of patients with recurrence of chronic GVHD after discontinuation of treatment? Does the risk of recurrent chronic GVHD decrease with time off treatment? If so, what is the minimum interval off treatment that would denote a population at little or no risk of recurrence of chronic GVHD?

Thank you,
Mara

Mara Miller, MA
Senior Regulatory Project Manager
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Hematology and Oncology Products
Division of Hematology Products
WO22, Room 2309
10903 New Hampshire Avenue
Silver Spring, MD 20993
(301) 796-0683 (phone)

Paul J. Martin, MD
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N, D2-100
P.O. Box 19204
Seattle, WA 98109-1024
Phone: (206)-667-4798
FAX: (206)-667-5155
e-mail: pmartin@fhcrc.org

January 20, 2015

Ann T. Farrell, MD
Food and Drug Administration
Center for Drug Evaluation and Research
White Oak Building 22, Room: 2309
10903 New Hampshire Avenue
Silver Spring, Maryland 20993

Re: PIND 124475

Dear Dr. Farrell,

This letter responds to correspondence received by e-mail on January 12, 2015 requesting additional information needed to prepare for our Type B meeting on January 28, 2015.

1. Corrected references for Supplementary Table 1 are provided on page 1 of the attached document.
2. Demographic characteristics, overall KM survival curves, and time to discontinuation of systemic treatment comparing first-line treatment with steroids alone versus steroids plus a calcineurin inhibitor are shown in Table 1 and Figure 1 on pages 2 – 4 of the attached document.
3. Demographic characteristics, overall KM survival curves, and time to discontinuation of systemic treatment comparing first salvage (second-line) treatment for patients refractory to steroids alone versus steroids plus a calcineurin inhibitor are shown on pages 5 – 7 of the attached document.
4. An extensive chart review would be needed to determine the percentage of patients with recurrence of GVHD after discontinuation of treatment. The available data clearly indicate that the risk of recurrent chronic GVHD decreases with time off treatment. Our best estimate is that approximately 5% of patients have recurrent GVHD more than 1 year after the first discontinuation of systemic treatment.

Please let us know if any additional information in advance of the meeting would assist your review. We look forward to a productive discussion with you and your colleagues on January 28, 2015.

Sincerely yours,

Paul J. Martin, MD
Member, Fred Hutchinson Cancer Research Center
Professor of Medicine, University of Washington

Correction of reference numbering in Supplementary Table 1. The corresponding references are provided below.

Supplementary Table 1. Randomized trials of treatment for chronic GVHD

Author	Arms compared	Double-blind	N	Results
Sullivan[9]	Initial prednisone ± azathioprine	Yes	179	Decreased survival
Koc[7]	Initial prednisone ± cyclosporine	No	287	Possible steroid-sparing effect
Koc[6]	Initial CSP/prednisone ± thalidomide	Yes	51	Toxicity
Arora[4]	Initial CSP/prednisone ± thalidomide	No	54	No benefit
Martin[8]	Initial CNI/prednisone ± MMF	Yes	151	No benefit
Gilman[5]	Initial CNI/prednisone ± hydroxychloroquine	No	54	Terminated early
Flowers[10] Greinix[11]	Second-line extracorporeal photopheresis	Yes	95	Possible steroid-sparing effect
Carpenter*	Initial sirolimus/prednisone ± CNI	No	151	Terminated early

*results not yet reported

4. Arora M, Wagner JE, Davies SM, et al. Randomized clinical trial of thalidomide, cyclosporine, and prednisone versus cyclosporine and prednisone as initial therapy for chronic graft- versus-host disease. *Biol Blood Marrow Transplant*. 2001;7(5):265-273.
5. Gilman AL, Schultz KR, Goldman FD, et al. Randomized trial of hydroxychloroquine for newly diagnosed chronic graft-versus-host disease in children: a Children's Oncology Group study. *Biol Blood Marrow Transplant*. 2012;18(1):84-91.
6. Koc S, Leisenring W, Flowers ME, et al. Thalidomide for treatment of patients with chronic graft-versus-host disease. *Blood*. 2000;96(12):3995-3996.
7. Koc S, Leisenring W, Flowers MED, et al. Therapy for chronic graft-versus-host disease: a randomized trial comparing cyclosporine plus prednisone versus prednisone alone. *Blood*. 2002;100:48-51.
8. 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.
9. Sullivan KM, Witherspoon RP, Storb R, et al. Prednisone and azathioprine compared with prednisone and placebo for treatment of chronic graft-v-host disease: prognostic influence of prolonged thrombocytopenia after allogeneic marrow transplantation. *Blood*. 1988;72(2):546-554.
10. Flowers ME, Apperley JF, van Besien K, et al. A multicenter prospective phase 2 randomized study of extracorporeal photopheresis for treatment of chronic graft-versus-host disease. *Blood*. 2008;112(7):2667-2674.
11. Greinix HT, van Besien K, Elmaagacli AH, et al. Progressive improvement in cutaneous and extracutaneous chronic graft-versus-host disease after a 24-week course of extracorporeal photopheresis--results of a crossover randomized study. *Biol Blood Marrow Transplant*. 2011;17(12):1775-1782.

Table 1. GVHD characteristics at onset of first-line treatment*

Characteristic	First-line treatment: Prednisone alone (n=33)	First-line treatment: Prednisone plus calcineurin inhibitor (n=170)
Median time from transplantation to first-line treatment, month (range)	**8.7 (4.0-57)†**	**6.5 (2.7-41)**
Sites involved, no. (%)		
Skin	24 (73)	131 (77)
Eyes	9 (27)	53 (31)
Mouth	29 (88)	153 (90)
Liver	19 (58)	84 (49)
Gastrointestinal tract		
Upper only	9 (27)	39 (23)
Any lower	3 (9)	32 (19)
Lung	**5 (15)**	**6 (4)**
Joint or fascia	4 (12)	19 (11)
Genital tract	1 (3)	14 (8)
No. of sites involved, no. (%)		
1 or 2	10 (30)	54 (32)
3	10 (30)	63 (37)
≥4	13 (39)	53 (31)
NIH global severity score, no. (%)		
Mild	0	0
Moderate	20 (61)	91 (54)
Severe	13 (39)	79 (46)
Subcategory of chronic GVHD		
Classic	6 (18)	13 (8)
Overlap	27 (82)	157 (92)
Karnofsky score, no. (%)		
80-100	**28 (85)**	**109 (64)**
<80	**5 (15)**	**61 (36)**

Platelet count, no. (%)		
<100 000/μL	**3 (9)**	**48 (28)**
≥100 000/μL	**30 (91)**	**122 (72)**
Serum total bilirubin, no. (%)		
>2 mg/dL	0	7 (4)
≤2 mg/dL	33 (100)	163 (96)
Progressive onset, no. (%)	0	6 (4)
Prior grade II-IV acute GVHD, no. (%)	**18 (55)**	**122 (72)**
Prednisone dose before starting treatment, no. (%)		
None	31 (94)	133 (78)
<0.5 mg/kg/day	2 (6)	30 (18)
0.5 - <1.0 mg/kg/day	0	6 (4)
≥1.0 mg/kg/day	0	1 (1)
Prednisone dose used for initial treatment, no. (%)		
<0.5 mg/kg/day	1 (3)	20 (12)
0.5 – <1.0 mg/kg/day	8 (24)	50 (29)
1.0 mg/kg/day	23 (70)	87 (51)
≥1.0 mg/kg/day	1 (3)	13 (7)

*****Data are adapted from the cohort reported by Inamoto et al. *Blood.* 2014;124(8):1363-1371 (Appendix VI in the briefing package).

†Notable differences are highlighted in bold

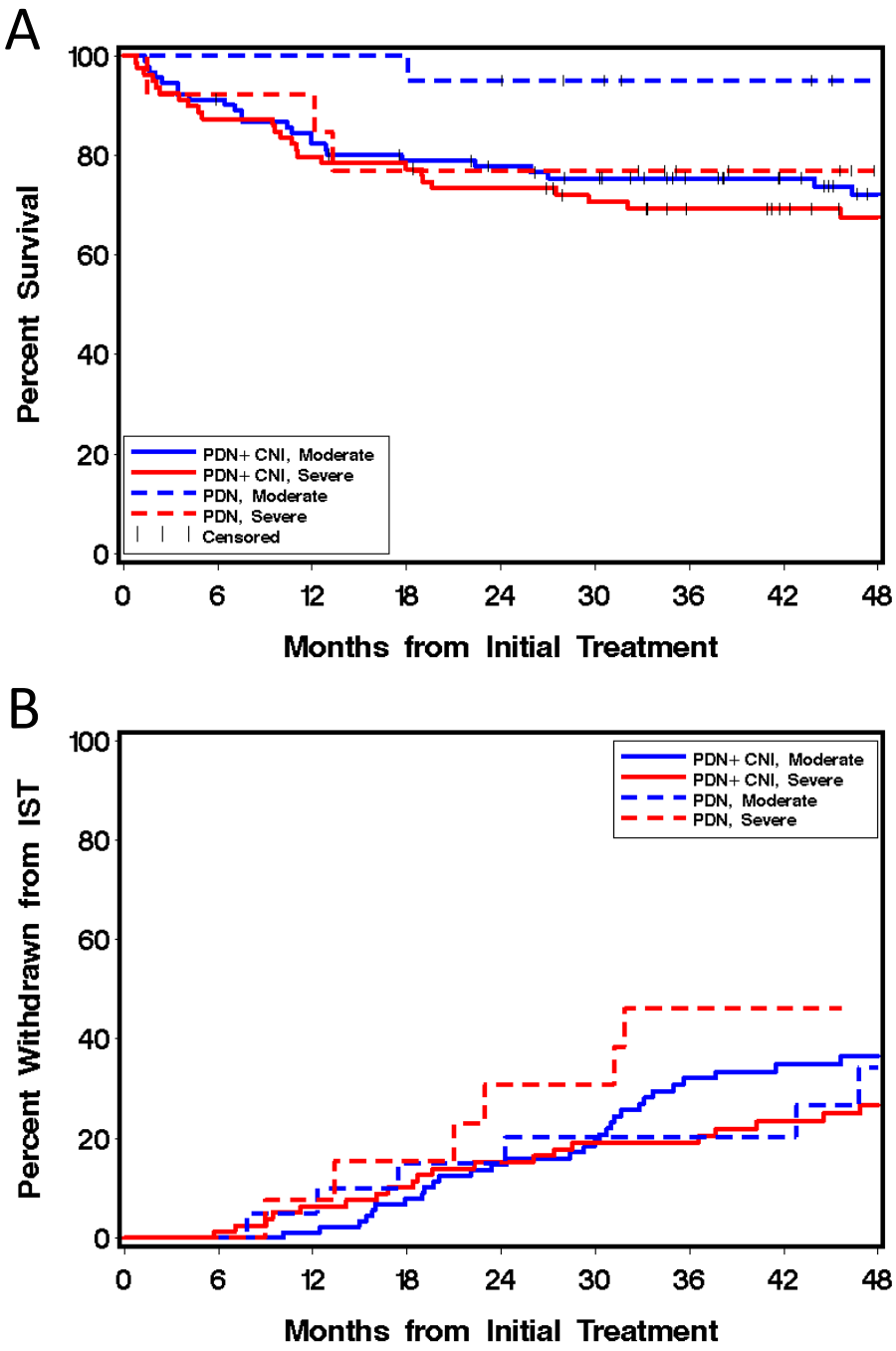


Figure 1. Outcomes after first-line therapy for chronic GVHD, according to treatment with prednisone (PDN) alone or prednisone plus a calcineurin inhibitor (CNI). Results for the cohort described in Table 1 are stratified according to NIH global severity score at the onset of treatment. Panel A shows survival, and panel B shows the cumulative incidence of withdrawal from systemic immunosuppressive treatment (IST) with death and recurrent malignancy as competing risks. Tic marks in panel A show end of follow-up.

Table 2. GVHD characteristics at onset of second-line systemic treatment, according to prior treatment*

	Prior Treatment	
	Prednisone alone (n=37)	**Prednisone plus calcineurin inhibitor (n=196)**
Median time from initial to second-line treatment, months (range)	**7.4 (0.3-61)†**	**3.9 (0.2-31)**
Reason for second-line treatment, no. (%)		
Progression	31 (84)	141 (72)
Lack of improvement	6 (16)	55 (28)
Sites involved, no. (%)		
Skin	29 (78)	139 (71)
Eyes	12 (32)	89 (45)
Mouth	27 (73)	141 (72)
Liver	11 (30)	57 (29)
Gastrointestinal tract		
Upper only	28 (14)	4 (11)
Any lower	28 (14)	1 (3)
Lung	8 (22)	22 (11)
Joint or fascia	10 (27)	49 (25)
Genital tract	4 (11)	20 (10)
Serosa	0	4 (2)
No. of sites involved, no. (%)		
1 or 2	14 (38)	72 (37)
3	14 (38)	61 (31)
≥4	9 (24)	63 (32)
NIH global severity score, no. (%)		
Mild	0	16 (8)
Moderate	19 (51)	102 (52)
Severe	18 (49)	78 (40)
Platelet count, no. (%)		
<100 000/μL	3 (8)	29 (15)
≥100 000/μL	34 (92)	163 (85)

Serum total bilirubin, no. (%)		
<2 mg/dL	32 (86)	185 (95)
≥2 mg/dL	5 (14)	10 (5)
Prednisone dose before second-line treatment, no. (%)		
None	**0**	**37 (19)**
<0.5 mg/kg/day	**18 (49)**	**88 (45)**
0.5 - <1.0 mg/kg/day	**11 (30)**	**41 (21)**
≥1.0 mg/kg/day	**8 (22)**	**29 (15)**
Second-line treatment, no. (%)		
Mycophenolate mofetil	**3 (8)**	**84 (43)**
Tacrolimus	**8 (22)**	**32 (16)**
Sirolimus	**4 (11)**	**31 (16)**
Extracorporeal photopheresis	**2 (5)**	**8 (4)**
Cyclosporine	**14 (38)**	**3 (2)**
Methotrexate	**2 (5)**	**8 (4)**
Other single agents	**2 (5)**	**13 (7)**
Multiple agents	**2 (5)**	**17 (9)**

*****Data are adapted from the cohort reported by Inamoto et al. *Blood*. 2013;121(12):2340-2346 (Appendix XII in the briefing package). Five patients were omitted from this analysis. One patient in the prednisone plus calcineurin inhibitor group was misclassified (should have been prednisone plus other), and NIH severity scores were missing in 4 other cases.

†Notable differences are highlighted in bold.

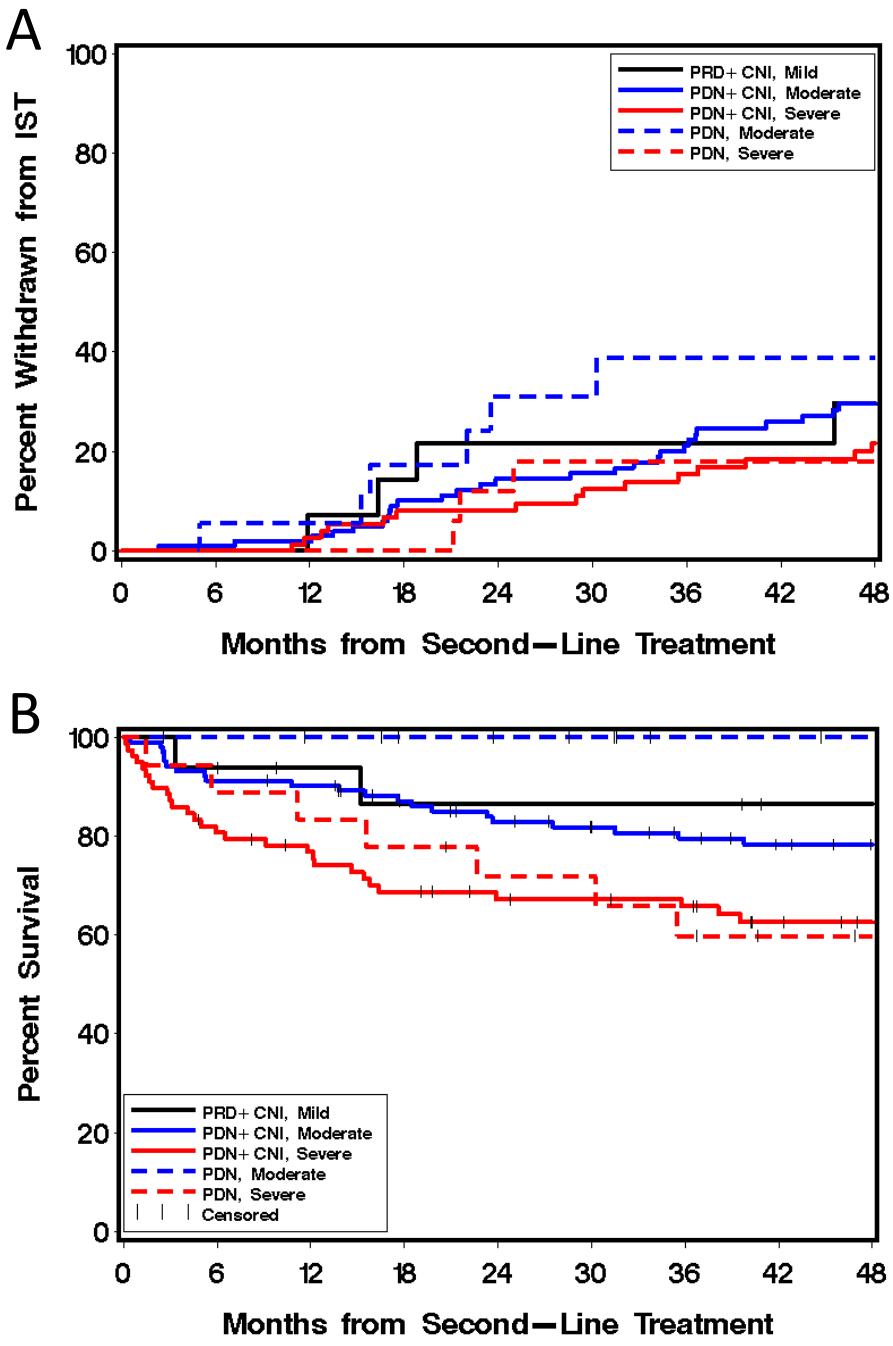


Figure 2. Outcomes after second-line (first salvage) therapy for chronic GVHD, according to prior treatment with prednisone alone or prednisone plus a calcineurin inhibitor (CNI). Results for the cohort described in Table 2 are stratified according to NIH global severity score at the onset of treatment. Panel A shows survival, and panel B shows the cumulative incidence of withdrawal from systemic immunosuppressive treatment (IST) with death and recurrent malignancy as competing risks. Tic marks in panel A show end of follow-up.

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

PIND 124475

MEETING MINUTES

Paul Martin, MD
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue, North, D2-100
P.O. Box 19204
Seattle, WA 98109-1024

Dear Dr. Martin:

Please refer to your Pre-Investigational New Drug Application (PIND) file for an Immunosuppressive agent.

We also refer to the meeting between representatives of your firm and the FDA on January 28, 2015. The purpose of the meeting was to discuss endpoints that could be used in the development path for products with indications related to the treatment of chronic GVHD.

A copy of the official minutes of the meeting is enclosed for your information. Please notify us of any significant differences in understanding regarding the meeting outcomes.

If you have any questions, call Mara Miller, Regulatory Project Manager at (301) 796-0683.

Sincerely,

{See appended electronic signature page}

Albert Deisseroth, MD, PhD
Clinical Team Lead
Division of Hematology Products
Office of Hematology and Oncology Products
Center for Drug Evaluation and Research

Enclosure:
Meeting Minutes

FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

MEMORANDUM OF MEETING MINUTES

Meeting Type: Type B
Meeting Category: Pre-IND

Meeting Date and Time: January 28, 2015
Meeting Location: White Oak Building #22, Room 1419

Application Number: PIND 124475
Indication: Chronic GVHD
Sponsor/Applicant Name: Paul Martin, MD

Meeting Chair: Albert Deisseroth, MD, PhD
Meeting Recorder: Mara Miller, MA

FDA ATTENDEES

Division of Hematology Products
Ann T. Farrell, MD, Director
Edvardas Kaminskas, MD, Deputy Director
Albert Deisseroth, MD, PhD, Clinical Team Lead
Donna Przepiorka, MD, PhD, Clinical Reviewer
Virginia E. Kwitkowski, MS, RN, ACNP-BC, Associate Director for Labeling (Acting)
Mara Miller, MA, Regulatory Project Manager

Office of Biostatistics
Kallappa Koti, PhD, Reviewer

Center for Biologics Evaluation and Review
Kristin Baird, MD, Reviewer

SPONSOR ATTENDEES

Paul J. Martin, MD, Member, Fred Hutchinson Cancer Research Center, Professor of Medicine, University of Washington
Corey Cutler, MD, Dana Farber Cancer Institute, Associate Professor of Medicine, Harvard University, by phone
Stephanie J. Lee, MD, Member, Fred Hutchinson Cancer Research Center, Professor of Medicine, University of Washington
Steven Z. Pavletic, MD, National Cancer Institute, Experimental Transplantation and Immunology Branch
Georgia Vogelsang, MD, Professor of Medicine (retired), Johns Hopkins University, by phone

Mary Horowitz, MD, Professor of Medicine, Medical College of Wisconsin
Sandra Mitchell, PhD, CRNP, National Cancer Institute, Outcomes Research Branch
John Koreth, MBBS, Dana Farber Cancer Institute, Associate Professor of Medicine, Harvard University, by phone
Jacqueline Mays, DDS, PhD, NIDCR, Intramural NIH Programs
Licia Masuch, Patient Care Coordinator, NIH, Experimental Transplantation and Immunology Branch

1.0 BACKGROUND

The purpose and objectives of this meeting were to identify endpoints that measure clinical benefit across different trial phases in the development of products indicated for treatment of chronic GVHD. No clinical development path has been mapped for indications related to treatment of chronic GVHD, and no products have been approved for such indications. Much of the current difficulty originates from the lack of a validated global clinical scoring system that could be used to measure response in studies intended for regulatory review. This pre-IND review provides an opportunity for a regulatory assessment of 5 potential clinical trial endpoints currently under active discussion in the academic community. These include failure-free survival, survival without progressive impairment, clinical response, patient reported outcomes, and an aggregate measure incorporating provider and patient assessments. Chronic GVHD investigators would benefit from better understanding of the extent to which these endpoints could be viewed as indicators of clinical benefit for purposes of regulatory review at each phase of the development path.

2. DISCUSSION

Question 1
To what extent could failure-free survival at 12 months be considered as an indicator of clinical benefit in early phase trials of initial systemic treatment for chronic GVHD?

FDA Response
We acknowledge that your evaluation of FFS showed that landmark analyses of this endpoint correlate with patient-reported outcomes, clinician-reported outcomes and the ability to discontinue systemic immunosuppression. However, we have a few concerns about use of FFS for regulatory purposes.

a) A landmark analysis of a time-to-event endpoint would not be interpretable in a single-arm trial, especially one that includes a population heterogeneous in prognostic factors.

Discussion
The FDA explained that landmark analysis refers to a point-in-time assessment (such as FFS at 12 months) as a binary endpoint.

b) FFS is a composite of efficacy and safety outcomes. With such an endpoint, it would not be possible to isolate a true measure of efficacy. If efficacy is demonstrated using an alternate

primary endpoint even in a randomized trial, FFS as you have defined is might be of use as an exploratory endpoint to confirm that there is no substantial adverse consequence with use of the new drug.

Discussion
The FDA hopes that the management of patients would have been developed sufficiently in Phase 2 to minimize toxicity in the pivotal trial. For a binary endpoint,those patients with toxicity and no response would be considered failures. In a time-to-event analysis they would be included, but an exploratory analysis with censoring for toxicity and a new therapy introduced should also be considered.

c) We agree that new treatment decisions are subject to bias. If you do proceed with this component, we recommend that you prespecify the criteria for when treatment should be changed, and use those criteria as the component instead.

Discussion
Using PFS requires using a population where lack of progression is truly beneficial and progression can be accurately assessed.

d) We recommend that you also confirm the correlations found in the landmark analyses in independent data sets.

Discussion
No discussion occurred.

Question 2
To what extent could failure-free survival at 6 months be considered as an indicator of clinical benefit in early phase trials of second-line systemic treatment for chronic GVHD?

FDA Response
See response to question 1.

Discussion
See discussion under question 1.

Question 3
Should the absence of recurrent or progressive malignancy be included as a component in the definition of failure-free survival?

FDA Response
See response to question 1.

Discussion
See discussion under question 1.

Question 4

Should steroid doses below a predefined threshold at 12 months after initial treatment or at 6 months after second-line treatment be included as an additional criterion of failure-free survival?

FDA Response

See response to question 1.

Discussion

See discussion under question 1.

Question 5

To what extent could prevention of "progressive impairment" at 2 years be considered as an indicator of clinical benefit in late phase trials of treatment for chronic GVHD?

- **Should any items be omitted from or added to the list of outcomes categorized as progressive impairment, or should any of the proposed item thresholds be modified?**
- **What additional analyses would help determine whether this endpoint represents clinical benefit?**
- **Given the typical time course of chronic GVHD, what would be the most appropriate time point for comparison between arms in a late-phase controlled trial using survival without progressive impairment as the primary endpoint?**
- **Given the typical time course of chronic GVHD, could earlier time points be used for comparison between arms in earlier-phase controlled trials using survival without progressive impairment as the primary endpoint?**

FDA Response

Regarding your proposal for use of a PFS endpoint, we have the following comments:

a) It is not clear from the briefing document how you define progression for this PFS endpoint. We suggest that the simplest approach would be a global measure of progression (such as is done for oncology trials) with prespecified subset analysis by organ system to ensure consistency.

Discussion

The FDA stated it depends on how PFS is defined. It will be up to the Sponsor to determine the components from Table 5 that are clinically meaningful. Patient input should be incorporated into the selection of these components. If a PRO tool is used, assessment of symptoms by the patient should not include signs or other determinations that would be best made by a clinician.

b) A time-to-event endpoint would not be interpretable in a single-arm trial, especially one that includes a population heterogeneous in prognostic factors. It would be helpful to use the single-arm trial to characterize this endpoint if it would be used as the primary outcome in a late-phase

randomized trial in the same study population. Such data from the early phase trial would be needed for sample size considerations.

Discussion
No discussion occurred.

c) The duration of follow-up needed for such a timepoint will depend on the characteristics determined in the earlier trial. We expect that the required follow-up may depend on the study population (i.e., shorter follow-up for populations that progress more quickly).

Discussion
No discussion occurred.

d) We caution that use of a PFS-like endpoint has inherent complications, including whether it represents a direct clinical benefit, missing data, frequency of assessment, criteria for progression and method of measurement, and potential bias, especially in an open-label trial. For additional information about these concerns, please see Section III.B.3.c in "Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics" at http://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf

Discussion
No discussion occurred.

e) Use of PFS as a clinical endpoint for a regulatory decision must be meaningful for the particular study population, which depends on its relevance to the direct clinical benefit, magnitude of the effect, and the risk-benefit of the new treatment compared to available therapies.

Discussion
No discussion occurred.

Question 6
Are changes in clinician-reported chronic GVHD manifestations sufficient to document clinical benefit?

FDA Response
You would need to show in an independent population that global response is at least reasonably likely to predict an improvement in survival. For example, in oncology durable response would appear to support activity of an agent in a population that is otherwise incurable. The general approach is to show an increase in response rate and that the responses are of a meaningful duration for the patient population. Data-based justification for use of CR +PR vs CR alone would be required. For patients expected to live long, response might be used as an objective measure of activity in conjunction with a PRO.

Discussion
FDA clarified the CR or CR+PR are examples of objective measures of activity.

Question 7
Could improvement in a patient-reported outcome tool be considered sufficient documentation of clinical benefit?

FDA Response
We agree that a PRO assessment would be useful in characterizing clinical benefit and might be acceptable as a key secondary or co-primary endpoint to measure the core disease-related symptoms of chronic GVHD. The PRO measure should be well-defined and reliable. A HRQoL measure may be included separately as an exploratory endpoint.
You have identified the Lee Symptom Scale and several organ symptom scales or global scales as patient-reported outcomes. Each has strengths and weaknesses. There is insufficient information in the briefing package to provide a complete review of these scales for regulatory purposes, and a discussion would require more time than allotted for this meeting. We encourage you to work closely with the Agency on specific PRO measure(s) proposed as key secondary or co-primary endpoints. We recommend that you request a separate meeting under PIND# 124475 to discuss the PRO measure(s).
In order to guide your development, selection, or modification of a well-defined and reliable PRO assessment intended to support labeling claims of treatment benefit, please refer to the principles specified in the FDA *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*, which can be found at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. It would be helpful if at least the items in Section III of this guidance were addressed in your briefing package for the meeting you request about a specific PRO measure for chronic GVHD trials.

Discussion
No discussion occurred.

Question 8
Would changes in a multicomponent clinical scale that incorporates clinician assessments, patient-reported outcomes and laboratory/functional measurements be sufficient to document changes in chronic GVHD disease activity?

FDA Response
From the limited description in your meeting package, it is unclear whether the clinician assessments, patient-reported outcomes, and laboratory/functional measurements have similar clinical importance and whether the results might be exclusively attributed to a subset of the components. In addition, it appears that you identified many of the components using an observational data of 497 patients because of their correlations with clinician-reported response. We recommend that you validate your results in independent datasets.
However, we suggest that you consider instead using simple endpoints wherever possible and pre-specify the other measures as additional secondary endpoints to test for internal consistency.

Discussion
FDA explained that a complex endpoint would certainly be acceptable if each component could be justified, but sample size considerations or patient population may warrant a simpler endpoint or co-primary endpoints instead.

Question 9
If yes, then what gold standard of clinical benefit should be used to develop and validate such a scale?

FDA Response
Clinician-reported outcomes and patient-reported outcomes that are well-defined and reliable in the intended population and context may be relevant measures of clinical benefit on their own. Whether a proposed response endpoint would need to be qualified against a "gold standard" would depend on the actual endpoint, the intended population and the context.

Discussion
No discussion occurred.

Additional Clinical Comments:

1. You may need to consider different endpoints for different patient populations. We might still expect an eventual demonstration of a survival benefit in a subgroup with relatively short OS, while for patients who live long but with the potential for disability, clinical response or a PRO might be more appropriate.

2. Although the ultimate goal is to prevent disability and induce tolerance, we caution that getting to that goal may take several steps. Endpoints may change over the course of time as new drugs alter the natural history of the disease.

3.0 ISSUES REQUIRING FURTHER DISCUSSION
None.

4.0 ACTION ITEMS
None.

5.0 ATTACHMENTS AND HANDOUTS
None.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

ALBERT B DEISSEROTH
01/29/2015