# MEASURING THERAPEUTIC RESPONSE IN CHRONIC GRAFT-VERSUS-HOST DISEASE

## NIH Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-Versus-Host Disease: IV. Response Criteria Working Group Report

**Correspondence:**

**Running Title:** Response criteria in chronic GVHD

**Keywords:** Chronic graft-versus-host disease, allogeneic cell transplantation, response criteria, consensus

**Word count:** abstract 153, text 6,528

**Tables:** 4

**Figures:** 2

## ABSTRACT

In 2005, the Response Criteria Working Group recommended several measures to document serial evaluation of chronic GVHD organ involvement. Provisional definitions of complete response, partial response, and progression were proposed for each organ and for overall outcome. Based on publications over the last eight years, the Working Group has updated its recommendations for measures and interpretation of organ and overall responses. Major changes included elimination of several clinical parameters from the determination of response, update or inclusion of specified organ scales to assess response, addition of the “mixed response” as an outcome category, and clarification of the need for protocol pre-specification of the role of irreversible changes and of local therapies in the assessment of response. Ancillary measures are strongly encouraged in clinical trials but are no longer recommended. Areas suggested for additional research include how to identify irreversible organ damage and validation of the modified response criteria, especially in the pediatric population.

## BACKGROUND

Overall survival or survival to permanent resolution of chronic GVHD and discontinuation of systemic immunosuppression are long-term clinical outcomes that are accepted as measures of meaningful benefit in chronic GVHD clinical trials,[1-3] but these long-term outcomes are not suitable for early-phase studies. Qualitative assessments of chronic GVHD manifestations can guide clinical decisions but are not adequate for measuring outcomes in clinical trials. To accelerate development of novel therapeutic agents in chronic GVHD, quantitative research tools are needed to measure short-term responses to treatment and to predict long-term clinical benefit.

The 2005 NIH recommendations proposed a broad set of assessment measures that were thought to be feasible in most academic settings and were based on group consensus with input from subspecialists. For this 2014 update, the reconvened Working Group reviewed the literature, and then used a consensus process to reconsider each prior recommendation while adding new recommendations. **Table 1** summarizes the 2014 changes to the original 2005 recommendations. Measures are designated as “recommended” if available data support their use for response measurement (**Table 2**), “strongly encouraged” if data are controversial or no alternate organ-specific measure exists, “exploratory” if the Working Group believes that substantial additional research is warranted before adoption, and “no longer recommended” if there are data supporting lack of usefulness or no additional data have been generated since the last Consensus Conference (**Table 3**).

## PURPOSE OF THIS DOCUMENT

This document summarizes proposed measures and criteria for use in clinical trials involving patients with chronic GVHD where the goal is to demonstrate improved patient outcomes or to obtain regulatory approval. The measures and criteria do not necessarily apply to routine patient care or to trials with limited resources or targeting specific organs. The following general principles were applied in selecting the recommended measures:

i. The measures should be easy for all care providers to use and should be available in the outpatient setting.

ii. The criteria should be adaptable for use in adults and in children.

iii. The measures should focus on the most important and common manifestations of chronic GVHD and should not attempt to characterize all possible clinical manifestations.

iv. Quantitative measures should be favored over qualitative measures.

v. Measurements of symptoms, global ratings, function, and quality of life should be made separately, and scales with established psychometric characteristics and desirable measurement properties should be used whenever possible.[4,5]

The Working Group had three additional goals: a) to propose provisional definitions of complete response, partial response, and disease progression for each organ and for overall response; b) to suggest appropriate strategies for using response measures in therapeutic clinical trials; and c) to outline future research directions.

**SUMMARY OF RECOMMENDATIONS**

1. "Chronic GVHD-specific" core measures include a) clinician- and patient-assessed signs and symptoms, b) the Lee Chronic GVHD Symptom Scale, and c) the clinician- or patient-reported global rating scales (**Table 2**).[6-8] Easily recorded continuous data should not be reduced to pre-specified categories.
2. "Chronic GVHD non-specific" ancillary measures for adults include either the SF-36 version 2 questionnaire[9,10] or the FACT-BMT[11] plus the Human Activity Profile (HAP) questionnaire[12] (**Table 3).** These measures are strongly encouraged but optional and should not be used as primary endpoints in chronic GVHD trials.
3. Age-appropriate modifications of existing measures should be used in children with chronic GVHD.[13-19]

4. Documenting response involves a comparison of chronic GVHD activity at two time points. Definitions of response are offered for each organ and for overall outcomes, although each protocol should define precisely how response will be determined. Simple forms to be used for clinician and patient assessments are provided in Appendices A and B, and at **insert website link here**. In each trial, irreversible baseline organ damage may be defined initially and then excluded from the requirements for achieving response. Currently, objective and subjective data are kept separate. The field would benefit from methods that integrate physical exam and laboratory findings with clinician and patient-reported information to accurately determine whether chronic GVHD is improving or worsening.
5. Measures should be made at regular intervals, for example every three months, and whenever a new systemic treatment is started or the patient stops study treatment. The minimum duration of response should be specified by the protocol, but in general should be at least 4 weeks.
6. Collaboration with sub-specialists is encouraged to develop more detailed organ- or site-specific measures that could improve the sensitivity of chronic GVHD assessments to change or serve as primary endpoints in organ-specific therapy trials. For example:
    a) Skin: skin-specific scoring systems,[20] durometer,[20-22] biopsy,[21] or imaging (ultrasound, MRI)[23,24]
    b) Eyes: corneal staining grading,[25,26] conjunctival grading,[27] Ocular Surface Disease Index[28]
    c) Oral: Oral Mucositis Rating Scale,[29] Oral Health Impact Profile-14,[30] saliva collection[31]
    d) Vulvar-vaginal: organ-specific grading[32,33]
    e) Function: range of motion measured by goniometer, fatigue severity scale[34-36]
7. Measures that predict outcomes, but are not sensitive to change or do not directly measure chronic GVHD activity should be collected at baseline but not used in the response assessment. Predictive measures to be recorded at baseline include performance status, platelet count, and the two-minute walk test.

## PROPOSED MEASURES OF CHRONIC GVHD RESPONSE ASSESSMENTS

The Working Group identified two broad categories of tools for use in the assessment of response. These include a) the "Chronic GVHD-specific" core measures that directly measure organ-specific manifestations of chronic GVHD, and b) the "non-specific" ancillary measures, which could reflect the overall impact of chronic GVHD, treatment, and other illness or co-morbidity on function or quality of life.

## CHRONIC GVHD-SPECIFIC CORE MEASURES

The core clinician-assessed and patient-reported chronic GVHD-specific measures are described in the following sections: organ-specific assessments, chronic GVHD symptoms, clinician- and patient-reported global ratings (**Table 2, Form A, Form B**). Specific pediatric considerations are highlighted where appropriate. For the assessment of symptoms in younger children, depending on the child's development, assistance can be provided by the health care provider or the parent. The Working Group also recommends formal, in-person training for all assessments in order to minimize intra- and inter-observer variability.[37-39] An instructional manual and slide set to assist with such training are available at **insert website link here**.

### Organ-specific assessments

The discussion below is applicable to signs and symptoms potentially attributable to chronic GVHD. If there is another documented reason for the abnormality, such as infection, injury, or other non-GVHD causes, it should be indicated on the case report forms, and the respective organ may not be evaluated for response assessment if so specified by the study protocol. The measures below reflect the minimum data capture in chronic GVHD trials. Some studies may require more detailed organ assessments that go beyond the minimum data elements.

*Skin and skin appendages.* Skin is the most frequently affected organ in chronic GVHD, and manifestations are highly variable. In the 2005 response criteria, proposed measures included the

percentage of body surface area (BSA) by the type of involvement (erythematous rash, movable sclerosis, non-moveable sclerosis). BSA measurement on a continuous scale suffered from poor inter-rater reliability, particularly for sclerosis.[38] Thus, the 2014 revision recommends the simpler NIH Skin Score instead, because it correlates with chronic GVHD severity, symptoms, and survival.[40] The NIH Skin Score is a 0-3 score that summarizes BSA involvement into four categories: no skin involvement, ≤18%, 19-50% and/or moveable sclerosis, and >50% and/or immovable sclerosis, impaired mobility, or ulcers. The "Rule of 9's" as an estimate of BSA involvement is intended for use in adults and is less accurate in children, particularly young children. For the sake of simplicity, we recommend using the "Rule of 9's" for all children, except for those less than one year of age. A body surface area grid for children less than 1 year of age can be found at **insert website link here** (Attachment 4). BSA assessment should include superficial skin eruptions, moveable sclerosis and immovable sclerosis. Superficial skin eruptions of chronic cutaneous GVHD include maculopapular, erythematous, lichen planus-like, papulosquamous, ichthyotic, and keratosis pilaris-like rashes. Superficial sclerosis (moveable) includes both lichen sclerosis-like and morphea-like lesions. Deep sclerosis includes diffuse, immovable (hidebound) sclerosis, fibrosis of subcutaneous fat septae ("rippling") and fasciitis ("groove sign"). The presence of ulcers should be noted but documentation of the size is no longer required.

Sclerotic changes are common in skin GVHD, difficult to measure reliably and respond slowly to therapy. In clinical trials of less than 6 months, sclerotic skin manifestations may be deemed "irreversible" at the beginning of the trial if little change during the trial period is anticipated. This may not be the case for other agents targeting sclerosis. Since quantitative methods to measure the depth of sclerotic involvement are not available in a general oncology practice, these changes have been described in more qualitative terms related to thickening, pliability, color, adherence to underlying tissues, or changes in joint mobility. No validated scale exists for assessing sclerotic skin changes of chronic GVHD. Measures such as the Rodnan scale for assessment of systemic sclerosis might be helpful for clinical evaluation, but this scale is not suitable for use in clinical trials because it does not

address the full spectrum of sclerotic skin manifestations in chronic GVHD. There is an urgent need for the development of more quantifiable and reproducible measurements or imaging methods that could be used in patients with sclerotic skin manifestations of chronic GVHD.[20-24] Alternatively, the 0-10 semi-quantitative scale for capturing clinician and patient-perceived severity might be adapted to collect perceptions of skin manifestations to help develop measurement methods where validated and practical measures do not currently exist. Changes in these scales may also be included in response assessments for therapeutic agents that are anticipated to impact skin sclerosis.

Pigmentary changes do not indicate activity in chronic GVHD disease *per se*. Moreover, changes in pigmentation occur gradually and are perceptible only across long time intervals. Thus, these changes are not scored for the purposes of response assessment.

Of the patient-reported measures of skin disease, the skin subscale of the Lee Symptom Scale correlates with severity of skin disease, and changes in patient-reported skin symptoms correlate with survival.[40] Patients should report their most severe itching during the past week, rated according to a 0 – 10 scale, since itching is the most frequent cutaneous symptom of chronic GVHD. These are considered recommended measures.

*Musculoskeletal connective tissue.* Assessment of joint range of motion is a very useful objective measure of chronic GVHD tissue response in patients with sclerotic changes involving large joints. The NIH Joint Score and the photographic range of motion scales correlate with change in joint involvement, and are recommended measures.[41]

*Eyes.* Dry eyes reflect either lacrimal dysfunction or destruction which may be permanent. Although the Schirmer's test[42] was recommended in the 2005 criteria, subsequent studies have not validated this test in ocular chronic GVHD, and it is no longer recommended as a response measure.[43] One study found that the NIH Eye Score, which scores patients from 0-3 on the basis of symptoms, need for eye drops, and use of therapeutic procedures or devices, could detect improvement or worsening in ocular chronic GVHD,[43] so this scale is a recommended measure.

Patients should report their “chief eye complaint” rated according to a 0 –10 scale for peak severity during the past week. The complaint can change from visit to visit, but only one “chief eye complaint” is graded. This method is simple to use but may impose undesirable limitations in patients with multiple complaints. The eye subscale of the Lee Symptom Scale and the Ocular Surface Disease Index are also sensitive to change,[43] but the eye subscale is more convenient since it is shorter and already included in the symptom battery.

*Mouth.* Previously, oral chronic GVHD was assessed using a modification of the Schubert Oral Mucositis Rating Scale (OMRS) that scores oral surfaces from 0 – 15, with higher scores indicating more severe involvement. The four chronic GVHD manifestations assessed in this scale included a) mucosal erythema based on the color intensity; b) lichen-type hyperkeratosis (percent of oral surface area); c) ulcerations (percent of oral surface area); and d) mucoceles (total number). Subsequent studies have suggested that mucoceles are not reliably assessed,[38,39] and their quantification does not correlate with important clinical outcomes.[44,45] Thus, the Working Group recommends removing mucoceles from the OMRS, resulting in a modified 0-12 scale. The term “hyperkeratosis” has been removed and clarified as “lichen-type changes” instead. Instructions for these assessments and a photo dictionary are provided in the instructional manual on the web: **insert website link here**.

Patients should report their mouth sensitivity (irritation resulting from normally tolerated spices, foods, liquids, or flavors), rated according to a 0 – 10 scale for peak severity during the past week. Children may have an easier time with a 0-3 scale, but this format has not been validated and may be less sensitive to change. Capture of mouth dryness and mouth pain on 0-10 scales is no longer recommended.[46] Mouth symptoms are also captured in the Lee Symptom Scale, a recommended measure.

*Hematopoietic.* Parameters to be captured at trial enrollment include platelet count[47] and absolute eosinophil count,[48,49] since they may have prognostic significance. However, hematologic values are not part of the response assessment, and their ongoing collection is no longer recommended.

*GI tract.* Gastrointestinal (GI) symptoms are difficult to measure quantitatively in the outpatient setting. GI symptoms during the preceding week are graded through interview by the examining clinician according to 0 – 3 severity scales for the upper and lower GI tract and esophagus, and GI tract overall. Patients with chronic GVHD often have weight loss that is not always explained by GI symptoms.[50] Although the exact relationship between weight loss and chronic GVHD activity is not clear, recording patient weight at each scheduled evaluation is strongly encouraged, given the simplicity of this measure and its potential importance for monitoring the success of therapy.

*Liver.* Liver injury should be assessed according to the most recent laboratory results for total serum bilirubin (mg/dL) and alanine aminotransferase (U/L). Laboratory upper limits of normal should also be recorded. Aspartate aminotransferase and alkaline phosphatase are not captured as they are not specific for liver inflammation.

*Lung.* The 2005 response criteria recommended the lung function score (LFS), based on forced expiratory volume in the first second (FEV1) and single breath diffusion lung capacity for carbon monoxide (DLCO) adjusted for hemoglobin, because it is predictive of respiratory failure and mortality after allogeneic hematopoietic stem cell transplantation.[51,52] The LFS has been used as a response measure in one trial in steroid-refractory patients suggesting sensitivity to change and utility as a response measure.[53] However, DLCO is not directly affected by bronchiolitis obliterans syndrome (BOS), and pulmonary function tests (PFTs) did not perform as well as the NIH Lung Symptom Score in predicting nonrelapse mortality and survival in an observational study, although 50% missing PFT data limited conclusions.[54] PFTs cannot be performed in children younger than 5 years, and DLCO usually cannot be measured in children less than 10 years old. FEV1 may not be reliable in children.[55] Another study that included mostly BOS patients showed that a decrease in FEV1 or FVC of greater than 10% was highly correlated with 5-year survival.[56]

The Working Group recommends recording the FEV1 (percentage of predicted value) and strongly encourages parallel capture of DLCO corrected for hemoglobin, forced vital capacity (FVC), total lung capacity (TLC), and residual volume (RV) to allow further validation studies, including

exploration of restrictive lung disease as a manifestation of chronic GVHD.[57-59] However, the age-adjusted FEV1 and the NIH Lung Symptom Score are currently recommended as primary response measures in BOS. If available, the FEV1 value should be prioritized first for response assessment. Exploration of the FEV1 slope as an outcome measure to account for the change in disease trajectory is encouraged.[60,61]

*Genitals.* Women should be asked specific questions relating to vulvar and vaginal symptoms, such as burning, pain, discomfort, or dyspareunia. Patients who report problems should be referred to a gynecologist. Since such symptoms could be under-reported or caused by conditions other than chronic GVHD,[62] and because proper evaluation requires a specialist exam, measures of genital response are considered exploratory. Both female and male genital symptoms may be captured by the exploratory item rating “Worst genital discomfort” on a scale from 0-10. Academic gynecologists interested in chronic GVHD are developing precise vulvo-vaginal assessment scales. These scales will be useful in selected trials where vulvar and vaginal changes are the primary endpoints of interest.[32,33]

Other organ systems may be affected by chronic GVHD, but are either rare or difficult to quantify by non-specialists. The Working Group encourages investigators to develop and validate response assessment tools that could detect meaningful clinical benefit in trials focused on specific organs or manifestations.

**Chronic GVHD symptoms**

Lee et al.[6] developed a symptom scale designed for individuals with chronic GVHD. The questionnaire asks patients to indicate the degree of *bother* that they experienced during the past four weeks due to symptoms in seven domains potentially affected by chronic GVHD (skin, eyes and mouth, breathing, eating and digestion, muscles and joints, energy, emotional distress). Published evidence supports its validity, reliability, and sensitivity to chronic GVHD severity.[40,43,54,63,64] The

symptom scale can be completed in approximately 5 minutes. The time frame may be changed to one week to capture more recent symptoms.

The Lee Symptom Scale has been tested only in patients older than 18 years. Given its face validity and other desirable properties, however, this scale could be used for assessment of chronic GVHD in pediatric patients using either child or parent report, after appropriate modification and psychometric evaluation.[65] Information may be obtained by self-report from adolescents over 12 years. For children who are 8-12 years of age, data should be obtained with the assistance of parents and the health care provider. Investigators are working on developing a symptom scale appropriate for all pediatric patients.[66]

The Lee Scale measures symptom bother as distinguished from symptom intensity, which is reported on the forms in Appendix B.[67] The degree to which patients report that they are bothered by a symptom represents a global assessment incorporating not only the intensity of the symptom and its frequency, but also the degree to which it causes emotional disturbance or interferes with functioning. The Lee Scale complements the information regarding the intensity and frequency of chronic GVHD symptoms. For example, oral sensitivity may be severe, but patients may report that they are not bothered or distressed by this symptom. By contrast, skin itching may not be very intense or frequent but may cause great distress. Additional investigators are needed to determine the relationships between symptom intensity, frequency, and distress or bother in patients with chronic GVHD and the degree to which these are distinct dimensions of the symptom experience.

**Clinician- and patient-reported global ratings**

*Clinician perceptions.* Physicians, nurse practitioners, or physician assistants should provide a subjective assessment of current overall chronic GVHD severity on a four-point scale (no chronic GVHD, mild, moderate, severe)[6] without knowledge of the calculated global severity score. They should also provide an assessment of current overall chronic GVHD severity on an 11-point numerical scale (0 indicates no GVHD manifestations; 10 indicates most severe chronic GVHD

symptoms possible).[68] The categories of mild, moderate, and severe have been used in previous studies for patient and clinician assessment, where they were often undefined but showed good prognostic characteristics.[6,69] Clinicians should also provide their assessments of chronic GVHD changes since the last assessment scored on a seven point scale (very much better, moderately better, a little better, about the same, a little worse, moderately worse, very much worse).[8] These semi-quantitative assessments may detect qualitative improvements that are clinically meaningful but not well captured using other measures. The protocol and data capture form should specify the time interval for assessing changes.

*Patient perceptions*. Similarly, at each patient self-assessment, patients should score their perceptions of overall chronic GVHD severity, overall severity of symptoms, and change in symptom severity compared with the last assessment, using the same response options used by clinicians.

The exact role of global scales in chronic GVHD response assessments and their appropriate utilization as outcome measures in clinical trials remain to be determined. These scales could be sensitive to qualitative changes that might otherwise escape detection if the assessments were limited to quantitative measures. They are used in studies that establish clinically meaningful changes in measures. A potential limitation is that personality traits can influence patient perceptions or self-report.[70]

## CHRONIC GVHD NONSPECIFIC ANCILLARY MEASURES

Non-specific measures of function and patient-reported outcomes related to functional status and health-related quality of life could potentially offer additive objective and subjective data regarding the effects of chronic GVHD and its therapy. The GVHD non-specific measures listed in **Table 3** assess different dimensions of the patient experience. These measures are strongly encouraged to allow investigation of the potential role of these non-specific measures as response measurements in chronic GVHD therapeutic clinical trials.

**Functional status**

For an extremely complex, multi-system disease such as chronic GVHD, objective measures of physical performance and patient-reported measures of functional status could represent important surrogate outcomes that might be more informative than the measures described above for assessing response in some situations (e.g., advanced skin sclerosis). At the very least, measures of functional status can provide corroborative evidence of important changes after therapy. In other patient populations with chronic diseases,[71-73] such outcomes have been extensively applied, and population norms for both physical performance measures and self-reported functional status are available. Since the use of functional endpoints in chronic GVHD assessment has not been extensively tested, and since these measures do not directly assess chronic GVHD manifestations, functional status outcomes can be used only as optional secondary endpoints in chronic GVHD trials until further information in available.

An objective measure of physical performance is the 2-minute walk distance (total distance in feet walked in 2 minutes)[74,75,76] measured at baseline as a prognostic factor since it correlates with survival,[77] but not as a response measure at follow-up. Although the measurement properties for the 2-minute walk distance have been less thoroughly examined than those of the 6-minute walk distance, the 2-minute walk may be a more feasible and efficient measure of performance in patients with chronic GVHD and impaired functional status, although it could suffer from a ceiling effect in the highly functional population. Whether the 6 min walk test would perform superior in the latter situation requires validation. Age-matched norms for walk time are available for adults and children. Grip strength[78-80] measured using a hydraulic dynamometer to capture muscle strength of the upper extremity[81] is no longer recommended as it does not correlate with chronic GVHD severity or outcome.[77]

*Human Activity Profile.* The strongly encouraged patient-reported measure of physical activity in adults is the Human Activity Profile (HAP) questionnaire. The 94 questions are ranked hierarchically in ascending order according to the metabolic equivalents of oxygen consumption required to perform

each activity.[12] The HAP therefore provides a survey of the activities 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. While the HAP correlates with chronic GVHD severity,[77,82] it may not be required if the SF36 is collected, since the SF36 also assesses physical activity. The Working Group no longer recommends the Activities Scale for Kids (ASK)[13-15] due to lack of data in patients with chronic GVHD.

*Performance scales.* The Karnofsky or Lansky Performance Scale is commonly used in clinical assessments of chronic GVHD and has prognostic value for survival, so it is strongly encouraged at enrollment.[83,84] It is not a valid measure of response.

**Self-Reported Health-Related Quality of Life**

The effects of chronic GVHD and its treatment on general physical and emotional health and quality of life are other patient-reported outcomes that may be responsive to change as a result of chronic GVHD therapy[85] although validation data are lacking so these instruments are only "strongly encouraged." The SF-36v2 (Medical Outcomes Study Short Form 36-item Questionnaire version 2) is a measure that 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.[10,86] The Functional Assessment of Chronic Illness Therapy (FACIT) is an oncology-specific quality-of-life instrument that has well-developed psychometric properties, and population norms for healthy individuals and those with both mild and more severe chronic illnesses. An additional 18-item disease-specific module evaluates concerns common to patients who have had hematopoietic cell transplantation (FACT-BMT).[11] These instruments are appropriate for adults, and in order to minimize burden for respondents, only one should be used. The SF-36v2 may be used alone. If the FACT-BMT is used, it should be combined with the HAP to capture functional abilities. A previously recommended pediatric scale, the Child Health Ratings Inventories (CHRIs) generic core and

Disease-Specific Impairment Inventory-HSCT,[17-19] is no longer recommended due to lack of data in chronic GVHD. Instead, the Working Group recommends exploration of the PedsQL for which there are more data in chronic illnesses.[87,88]

Cross-sectional studies have shown that chronic GVHD has an adverse effect on quality of life,[89] but the role of quality of life as a measure of response to therapy or as a predictor of long-term outcome remains to be defined.[90] Patient-reported quality-of-life measures can augment but cannot replace quantitative measures of chronic GVHD activity in clinical trials. Since responses to patient surveys may be affected by personality traits[70] and baseline status, cross-sectional measurements may not be interpretable.

## CHRONIC GVHD DATA COLLECTION FORMS

Appendices A and B at: **insert website link here** (Forms A and B) provide downloadable data collection forms for the recommended clinician-assessed and patient-reported measures. In clinical trials, data should be submitted to the study-coordinating center for further calculations, processing, and interpretation of responses. It is not necessary to include all recommended or strongly encouraged measures in every trial, and judgment must be used in deciding which items will best suit the needs of each study. In all studies, the measures to be collected and the timing of the assessments must be specified.

## PROVISIONAL CRITERIA FOR DEFINITION OF RESPONSE

In order to assess response, disease manifestations at two time points must be compared, and a judgment must be made as to whether the magnitude of any change qualifies as clinical improvement or clinical deterioration. The magnitude of change required for clinical improvement or deterioration should reflect genuine clinical change, and the criteria should be developed and standardized as much as possible to avoid measurement error. This standardization may be relatively easy to establish for manifestations that can be measured quantitatively with little day-to-day variation but will

be more difficult to establish for manifestations that can be measured only in more qualitative ways. The Working Group proposes the following consensus definitions for assessment of overall response and for measurable response within an organ (**Table 4**). Proposed guidelines for calculating responses, an online calculator, and instructions for use by study coordinating centers are also available on the web at **insert website link here**.

**Objective measures of GVHD activity**

*Overall response.* Three general categories of overall response are proposed for interpretation of clinical trials: complete response, partial response, and other (unchanged, mixed response, progression). Complete overall response is defined as resolution of all reversible manifestations in each organ or site, and partial overall response is defined as improvement in a measure for at least one organ or site without progression in measures for any other organ or site as described in the following sections on organ response. The Working Group recommends that skin, mouth, liver, upper and lower GI, esophagus, lung, eye, and joint/fascia be considered in evaluating overall response. Genital tract and other manifestations are not included due to lack of validated response measures.

*Complete organ response*. The term “complete organ response” indicates resolution of all reversible manifestations related to chronic GVHD in a specific organ. This category may not apply to organs with irreversible damage. For patients with BOS, a partial response or stable disease, as measured by FEV1, combined with a complete response in all other organs can be considered an overall complete response.

*Partial organ response.* The proposed general guideline for defining partial response in a specific organ requires a change in score from baseline that reflects genuine clinical benefit and exceeds the measurement error of the assessment tool: an improvement of 1 or more points on a 4-7 point scale, or an improvement of 2 or more points on a 10-12 point scale. Partial response in the liver requires at least a 50% improvement in the ALT or AST if the baseline is 3 or more times the upper limit of normal (ULN) or the total bilirubin is greater than 3 mg/dL. If the baseline values for ALT or AST are

less than 3x the ULN or total bilirubin is less than or equal to 3 mg/dL, then only normalization of values (a complete response) is considered a response. For patients with BOS, an absolute improvement in FEV1 of 10% or more (e.g., 50% to 60%) is considered a partial response[56] while normalization is considered a complete response.

Because the NIH 0-3 Skin Score is now recommended for response assessment, even substantial improvement in sclerotic features will not be considered responses unless the NIH Skin Score improves. Trials targeting sclerotic chronic GVHD should use more detailed measures to document change in extent or functional consequences of sclerosis.

*Organ Progression.* Criteria for progression in each organ must be defined, since the overall category of partial response requires the absence of progression in any organ (see below). For skin, mouth, and the GI tract, a worsening of 1 point or more on a 4-7 point scale or a worsening of 2 or more points on a 10-12 point scale indicates progression, whether or not the organ was previously involved. An exception is a change in the lower GI score from 0 to 1 which is not counted as progression, since many patients have mild, intermittent, self-limited diarrhea, and one study showed poor test-retest reliability in 0 to 1 changes.[91] Worsening of liver GVHD is defined based on the baseline value. If the baseline ALT or AST is less than 3 times the ULN or total bilirubin is 3 mg/dL or less then worsening by 2x ULN or more is considered progression. If the baseline ALT or AST is greater than 3 times the ULN or total bilirubin is more than 3 mg/dL, then worsening by 3x ULN or more is considered progression. For patients with BOS, worsening of absolute FEV1 by 10% or more (e.g., 50% to 40%) is considered progression. Progression cannot be scored for manifestations with baseline values that are too close to the worst score.

The Working Group noted that “trivial” progression can occur, where new organ involvement occurs or organ involvement is worse as measured by the scale but the change in score does not reach the threshold defined as clinically meaningful (the standard of care would not dictate a change of therapy, and no new functional limitations are present). These patients should not be classified as

having progressed; they will fall into the unchanged category. Provision for such trivial changes should be defined in the protocol.

*Mixed Response.* Mixed response is a new category defined as complete or partial response in one or more organs accompanied by progression in one or more organs. This category should be considered progression for the purposes of analysis but may aid in identifying organ-specific response patterns.

*Unchanged.* Patients who do not meet the criteria for complete response, partial response, progression or mixed response are considered unchanged. Unchanged patients will generally be considered nonresponders unless specified otherwise by the protocol.

*Limitations in measurement of organ responses*. The response criteria do not account for qualitative changes. Clinical experience indicates that clinically important qualitative improvement often occurs before improvement in the objective measures. For this reason, the response criteria are not intended for use as the primary guide for clinical decisions. Certain organs or rare manifestations are not considered in the response criteria because quantitative assessments are not feasible but may be the most important manifestations of chronic GVHD for individual patients. To capture qualitative and global changes of the entire chronic GVHD syndrome, use of the clinician and patient-assessed 0-10 global and 7 point change scales is strongly encouraged. The response criteria also do not account for the prior trajectory of abnormalities. For example, “stable” or “unchanged” disease might be considered a meaningful response when the prior trajectory was clear progression, as indicated, for example, by serial pulmonary function tests or rapidly progressive sclerosis, whereas “stable disease” after prior improvement or stability should not be considered a “response”. Data captured to record the disease trajectory prior to enrollment should be specified by the protocol.

*Validation of response criteria.* One study showed significant prognostic value of the NIH calculated responses for predicting survival in the context of a therapeutic trial[53] but another observational study did not.[92] The criteria proposed in these guidelines have been modified based on

publications since 2005 but still need to be validated prospectively for patients with chronic GVHD. For these reasons, the updated criteria are still provisional and subject to change with further clinical experience. Also, depending on the stringency of response definitions required by the specific study, these general guidelines could be modified to fit the needs of a particular protocol. Since the criteria are subject to change, we strongly recommend that data report forms should always record the actual numerical values for any measurement.

*Caveats for assessment of response in clinical trials.* Protocols must specify the times when response will be assessed, and the requirement(s) for durability of response. The recommended minimum duration of response is 4 weeks (see 2014 “Design of Clinical Trials” Working Group report). Permanent discontinuation of systemic chronic GVHD therapy confirmed for the duration of the observation period indicates a durable response. If additional therapy for chronic GVHD is added before the end of the study period, the patient is not evaluable for response and is considered a treatment failure in that organ (if organ-directed therapy was added) or overall (if systemic therapy was added).

For certain organs and measures, chronic GVHD sequelae can reflect damage that is not reversible. For these manifestations, protocols may specify that determination of overall complete response may ignore these residual abnormalities. The progression category applies to all organs. Certain manifestations such as dry eyes, dry mouth, esophageal stricture, bronchiolitis obliterans, or advanced sclerotic skin lesions may be designated as irreversible and may be excluded from consideration for assessments of overall complete or partial response, if specified by the protocol. Addition of topical or organ-directed treatments for the eyes, mouth, esophagus and genital tract generally make it impossible to assess the response to systemic treatment. For example, if esophageal dilation is performed, the esophagus is not evaluable for response. The use of therapeutic eye procedures such as punctal plugging or ligation, use of devices such as scleral or bandage lenses, or addition of topical ophthalmic treatments during a clinical trial may make the eye not evaluable. Some topical therapies may result in systemic effects due to absorption (e.g., GI tract,

lung, skin), and protocols should specify whether such treatments make these organs or overall response not evaluable.

**Subjective measures of GVHD activity**

While improvement in patient functioning and symptoms is recognized as a measure of clinical benefit, the Working Group recommends that clinician-reported and patient-reported outcomes be tabulated separately from objective responses for now. The terms “complete response,” “partial response,” and “progression” do not technically apply to subjective or functional measures data. Clinician- and patient-reported outcomes should be classified into response (clinically meaningful improvement) versus no response (no improvement or worsening), as measured by change between baseline and follow-up scores. The definition of improvement or worsening for such scales is based on the reliability of the measure (the variability due to measurement error) and is anchored against clinically perceptible changes. For global ratings and categorical scales, a 1-point change on a 0-3 or 1-7 point scale, or a 2 – 3-point change (0.5 standard deviation change) on a 0 – 10-point scale could be considered clinically meaningful.

Unless otherwise specified, for all patient-reported measures, a change of 0.5 standard deviation may be considered clinically meaningful for normally distributed data.[93,94] For example, a distribution-based analysis was used to define improvement as a change of 6 – 7 points (0.5 SD) on the total chronic GVHD symptom score.[6] For the physical and mental component summary scores for the SF36, a change of 5 points is considered clinically meaningful.[95,96] For HAP, clinically meaningful improvement is defined as a 10-point increase in the maximum activity score, since a change of this magnitude is sufficient to change the disability category at the middle of the scale.

An area for future investigation is to determine methods to integrate objective and subjective measures into holistic assessments of chronic GVHD disease activity. This approach has been used in other autoimmune diseases, for example, the American College of Rheumatology criteria for rheumatoid arthritis,[97,98] Crohn’s Disease Activity Index (CDAI),[99,100] ankylosing spondylitis short-term

improvement criteria,[101] the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K),[102-105] Systemic Lupus Erythematosis Responder Index,[106] and the British Isles Lupus Assessment Group (BILAG).[107,108]

## UTILIZATION OF RESPONSE ASSESSMENT AS A PRIMARY ENDPOINT IN CLINICAL TRIALS

The use of the recommended measures and methods of calculating response has at least two important advantages. First, the field will gain valuable experience with the measures to help validate them in the context of therapeutic clinical trials. Second, investigators will be using standardized assessment methods that will allow comparison of efficacy between agents, and allow recalculation of response categories if new information becomes available. It is imperative that the field identify valid methods of documenting whether a patient is responding adequately to treatment for chronic GVHD.

More sophisticated assessments of certain organs such as skin, eyes, mouth, female genital tract, and joints may be needed for certain studies.[20-25,27-29,32,33] Specialized expertise will be needed for these assessments, and the criteria for measurement of response in these situations exceed the scope of the current proposal. The Working Group encourages development and validation of more precise assessment tools that could be used in organ-specific trials. In situations where expert assessors are not readily available, objective assessments of the skin, mouth, eye and external genitalia might be enhanced through review of serial photographs by a panel of expert individuals as blinded assessors who have no other information about the patient, so as to avoid potential inter-rater differences.[39]

Note that this document addresses only the measurement of clinical responses. It does not address other suggested surrogate endpoints such as failure-free survival, defined as absence of relapse, death and addition of new treatment,[109] that do not rely on direct assessment of organ responses.

## FUTURE DIRECTIONS

The proposed response criteria are expected to enhance uniformity and feasibility of data collection methods and further advance standards of chronic GVHD clinical trials. Although this 2014 proposal is based on substantial interim evidence of utility and suggested clinical benefit for many proposed measures, these recommendations need to be tested further in prospective chronic GVHD therapy trials. Developing algorithms for measuring combinations of organ specific and global assessment responses to assess overall therapy response, and definition of minimal clinically meaningful cut off points, similar to what has been successfully done in other systemic inflammatory disease, may help in the development of highly relevant cGvHD response measurement tools. Improved methods will be needed to distinguish chronic GVHD disease activity from irreversible damage and to develop a chronic GVHD activity index for clinical trials, perhaps enhanced through the use of biomarkers.[110]

## References


1. Stewart BL, Storer B, Storek J, et al: Duration of immunosuppressive treatment for chronic graft-versus-host disease. Blood 104:3501-6, 2004
2. Pavletic SZ, Carter SL, Kernan NA, et al: Influence of T-cell depletion on chronic graft-versus-host disease: results of a multicenter randomized trial in unrelated marrow donor transplantation. Blood 106:3308-13, 2005
3. Koc S, Leisenring W, Flowers ME, et al: Therapy for chronic graft-versus-host disease: a randomized trial comparing cyclosporine plus prednisone versus prednisone alone. Blood 100:48-51, 2002
4. Acquadro C, Berzon R, Dubois D, et al: Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health 6:522-31, 2003
5. Revicki DA, Osoba D, Fairclough D, et al: Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 9:887-900, 2000
6. Lee S, Cook EF, Soiffer R, et al: Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. Biol Blood Marrow Transplant 8:444-52, 2002
7. Cleeland CS, Mendoza TR, Wang XS, et al: Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. Cancer 89:1634-46, 2000
8. Osoba D, Rodrigues G, Myles J, et al: Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 16:139-44, 1998
9. Ware JE, Jr., Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473-83, 1992
10. Ware JE, Jr.: SF-36 health survey update. Spine 25:3130-9, 2000
11. 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 19:357-68, 1997
12. Daughton DM, Fix AJ, Kass I, et al: Maximum oxygen consumption and the ADAPT quality-of-life scale. Arch Phys Med Rehabil 63:620-2, 1982
13. Plint AC, Gaboury I, Owen J, et al: Activities scale for kids: an analysis of normals. J Pediatr Orthop 23:788-90, 2003
14. Young NL, Williams JI, Yoshida KK, et al: Measurement properties of the activities scale for kids. J Clin Epidemiol 53:125-37, 2000
15. Young NL, Yoshida KK, Williams JI, et al: The role of children in reporting their physical disability. Arch Phys Med Rehabil 76:913-8, 1995
16. Lansky SB, List MA, Lansky LL, et al: The measurement of performance in childhood cancer patients. Cancer 60:1651-6, 1987
17. Parsons SK, Barlow SE, Levy SL, et al: Health-related quality of life in pediatric bone marrow transplant survivors: according to whom? Int J Cancer Suppl 12:46-51, 1999
18. Parsons SK, Shih MC, Duhamel KN, et al: Original Research Article: Maternal Perspectives on Children's Health-Related Quality of Life During the First Year After Pediatric Hematopoietic Stem Cell Transplant. J Pediatr Psychol, 2005
19. Parsons SK, Shih MC, Mayer DK, et al: Preliminary psychometric evaluation of the Child Health Ratings Inventory (CHRIs) and Disease-Specific Impairment Inventory-Hematopoietic Stem Cell Transplantation (DSII-HSCT) in parents and children. Qual Life Res 14:1613-25, 2005
20. Seyger MM, van den Hoogen FH, de Boo T, et al: Reliability of two methods to assess morphea: skin scoring and the use of a durometer. J Am Acad Dermatol 37:793-6, 1997
21. Aghassi D, Monoson T, Braverman I: Reproducible measurements to quantify cutaneous involvement in scleroderma. Arch Dermatol 131:1160-6, 1995
22. Falanga V, Bucalo B: Use of a durometer to assess skin hardness. J Am Acad Dermatol 29:47-51, 1993

23. Gottlober P, Leiter U, Friedrich W, et al: Chronic cutaneous sclerodermoid graft-versus-host disease: evaluation by 20-MHz sonography. J Eur Acad Dermatol Venereol 17:402-7, 2003
24. Dumford K, Anderson JC: CT and MRI findings in sclerodermatous chronic graft vs. host disease. Clin Imaging 25:138-40, 2001
25. Bron AJ, Evans VE, Smith JA: Grading of corneal and conjunctival staining in the context of other dry eye tests. Cornea 22:640-50, 2003
26. Ogawa Y, Kim SK, Dana R, et al: International Chronic Ocular Graft-vs-Host-Disease (GVHD) Consensus Group: proposed diagnostic criteria for chronic GVHD (Part I). Sci Rep 3:3419, 2013
27. Robinson MR, Lee SS, Rubin BI, et al: Topical corticosteroid therapy for cicatricial conjunctivitis associated with chronic graft-versus-host disease. Bone Marrow Transplant, 2004
28. Schiffman RM, Christianson MD, Jacobsen G, et al: Reliability and validity of the Ocular Surface Disease Index. Arch Ophthalmol 118:615-21, 2000
29. Schubert MM, Williams BE, Lloid ME, et al: Clinical assessment scale for the rating of oral mucosal changes associated with bone marrow transplantation. Development of an oral mucositis index. Cancer 69:2469-77, 1992
30. Slade GD: Assessing change in quality of life using the Oral Health Impact Profile. Community Dent Oral Epidemiol 26:52-61, 1998
31. Imanguli MM, Atkinson JC, Mitchell SA, et al: Salivary gland involvement in chronic graft-versus-host disease: prevalence, clinical significance, and recommendations for evaluation. Biol Blood Marrow Transplant 16:1362-9, 2010
32. Spinelli S, Chiodi S, Costantini S, et al: Female genital tract graft-versus-host disease following allogeneic bone marrow transplantation. Haematologica 88:1163-8, 2003
33. Stratton P TM, Turner M.: Vulvar and vaginal graft versus host disease in women after hematopoietic stem cell transplantation. J Soc Gynecol Invest 11:162A, 2004
34. Hann DM, Denniston MM, Baker F: Measurement of fatigue in cancer patients: further validation of the Fatigue Symptom Inventory. Qual Life Res 9:847-54, 2000
35. Buysse DJ, Reynolds CF, 3rd, Monk TH, et al: The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 28:193-213, 1989
36. Carpenter JS, Andrykowski MA: Psychometric evaluation of the Pittsburgh Sleep Quality Index. J Psychosom Res 45:5-13, 1998
37. Greinix HT, Pohlreich D, Maalouf J, et al: A single-center pilot validation study of a new chronic GVHD skin scoring system. Biol Blood Marrow Transplant 13:715-23, 2007
38. 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
39. Treister NS, Stevenson K, Kim H, et al: Oral chronic graft-versus-host disease scoring using the NIH consensus criteria. Biol Blood Marrow Transplant 16:108-14, 2010
40. 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 120:2545-52; quiz 2774, 2012
41. Inamoto Y, Pidala J, Chai X, et al: Joint and fascia manifestations in chronic graft-versus-host disease and their assessment. Arthritis Rheum 66(4):1044-1052, 2014
42. Vitali C, Bombardieri S, Jonsson R, et al: Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis 61:554-8, 2002
43. Inamoto Y, Chai X, Kurland BF, et al: Validation of measurement scales in ocular graft-versus-host disease. Ophthalmology 119:487-93, 2012
44. Elad S, Zeevi I, Or R, et al: Validation of the National Institutes of Health (NIH) scale for oral chronic graft-versus-host disease (cGVHD). Biol Blood Marrow Transplant 16:62-9, 2010
45. 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 49:116-21, 2014

46. Fassil H, Bassim CW, Mays J, et al: Oral chronic graft-vs.-host disease characterization using the NIH scale. J Dent Res 91:45S-51S, 2012
47. Lee SJ, Vogelsang G, Flowers ME: Chronic graft-versus-host disease. Biol Blood Marrow Transplant 9:215-33, 2003
48. Jacobsohn DA, Schechter T, Seshadri R, et al: Eosinophilia correlates with the presence or development of chronic graft-versus-host disease in children. Transplantation 77:1096-100, 2004
49. Baird K, Steinberg SM, Grkovic L, et al: National Institutes of Health chronic graft-versus-host disease staging in severely affected patients: organ and global scoring correlate with established indicators of disease severity and prognosis. Biol Blood Marrow Transplant 19:632-9, 2013
50. Jacobsohn DA, Margolis J, Doherty J, et al: Weight loss and malnutrition in patients with chronic graft-versus-host disease. Bone Marrow Transplant 29:231-6., 2002
51. Parimon T, Madtes DK, Au DH, et al: Pretransplant lung function, respiratory failure, and mortality after stem cell transplantation. Am J Respir Crit Care Med 172:384-90, 2005
52. Walter EC, Orozco-Levi M, Ramirez-Sarmiento A, et al: Lung function and long-term complications after allogeneic hematopoietic cell transplant. Biol Blood Marrow Transplant 16:53-61, 2010
53. 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
54. 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 20:337-44, 2014
55. Kalhoff H, Breidenbach R, Smith HJ, et al: Spirometry in preschool children: time has come for new reference values. J Physiol Pharmacol 60 Suppl 5:67-70, 2009
56. 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 18:1044-54, 2012
57. Bergeron A, Bengoufa D, Feuillet S, et al: The spectrum of lung involvement in collagen vascular-like diseases following allogeneic hematopoietic stem cell transplantation: report of 6 cases and review of the literature. Medicine (Baltimore) 90:146-57, 2011
58. Freudenberger TD, Madtes DK, Curtis JR, et al: Association between acute and chronic graft-versus-host disease and bronchiolitis obliterans organizing pneumonia in recipients of hematopoietic stem cell transplants. Blood 102:3822-8, 2003
59. Yoshihara S, Yanik G, Cooke KR, et al: Bronchiolitis obliterans syndrome (BOS), bronchiolitis obliterans organizing pneumonia (BOOP), and other late-onset noninfectious pulmonary complications following allogeneic hematopoietic stem cell transplantation. Biol Blood Marrow Transplant 13:749-59, 2007
60. Anthonisen NR, Connett JE, Kiley JP, et al: Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. JAMA 272:1497-505, 1994
61. Scanlon PD, Connett JE, Waller LA, et al: Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung Health Study. Am J Respir Crit Care Med 161:381-90, 2000
62. Shanis D, Merideth M, Pulanic TK, et al: Female long-term survivors after allogeneic hematopoietic stem cell transplantation: evaluation and management. Semin Hematol 49:83-93, 2012
63. Mitchell SA, Leidy NK, Mooney KH, et al: Determinants of functional performance in long-term survivors of allogeneic hematopoietic stem cell transplantation with chronic graft-versus-host disease (cGVHD). Bone Marrow Transplant 45:762-9, 2010
64. Treister N, Chai X, Kurland B, et al: Measurement of oral chronic GVHD: results from the Chronic GVHD Consortium. Bone Marrow Transplant, 2013
65. Matza LS, Swensen AR, Flood EM, et al: Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. Value Health 7:79-92, 2004
66. Wiener L, Baird K, Crum C, et al: Child and parent perspectives of the chronic graft-versus-host disease (cGVHD) symptom experience: a concept elicitation study. Support Care Cancer 22:295-305, 2014

67. Goodell TT, Nail LM: Operationalizing symptom distress in adults with cancer: a literature synthesis. Oncol Nurs Forum 32:E42-7, 2005
68. Preston CC, Colman AM: Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst) 104:1-15, 2000
69. Lee SJ, Klein JP, Barrett AJ, et al: Severity of chronic graft-versus-host disease: association with treatment-related mortality and relapse. Blood 100:406-14, 2002
70. Herzberg P, Lee SJ, Heussner P, et al: Personality influences quality of life assessments in patients after allogeneic hematopoietic stem cell transplantation - results from a joint evaluation of the prospective German multi-center validation trial and the Fred Hutchinson Cancer Research Center. Bone Marrow Transplant, in press, 2012
71. Nagashima M, Shu G, Yamamoto K, et al: The ability of disease modifying antirheumatic drugs to induce and maintain improvement in patients with rheumatoid arthritis. epidemiology of DMARDs treatment in Japan. Clin Exp Rheumatol 23:27-35, 2005
72. Koller WC, Lyons KE, Truly W: Effect of levodopa treatment for parkinsonism in welders: A double-blind study. Neurology 62:730-3, 2004
73. Craig J, Young CA, Ennis M, et al: A randomised controlled trial comparing rehabilitation against standard therapy in multiple sclerosis patients receiving intravenous steroid treatment. J Neurol Neurosurg Psychiatry 74:1225-30, 2003
74. Waters RL, Lunsford BR, Perry J, et al: Energy-speed relationship of walking: standard tables. J Orthop Res 6:215-22, 1988
75. Brooks D, Parsons J, Tran D, et al: The two-minute walk test as a measure of functional capacity in cardiac surgery patients. Arch Phys Med Rehabil 85:1525-30, 2004
76. Eiser N, Willsher D, Dore CJ: Reliability, repeatability and sensitivity to change of externally and self-paced walking tests in COPD patients. Respir Med 97:407-14, 2003
77. Pidala J, Chai X, Martin P, et al: Hand grip strength and 2-minute walk test in chronic graft-versus-host disease assessment: analysis from the Chronic GVHD Consortium. Biol Blood Marrow Transplant 19:967-72, 2013
78. Mathiowetz V, Kashman N, Volland G, et al: Grip and pinch strength: normative data for adults. Arch Phys Med Rehabil 66:69-74, 1985
79. Mathiowetz V, Weber K, Volland G, et al: Reliability and validity of grip and pinch strength evaluations. J Hand Surg [Am] 9:222-6, 1984
80. Mathiowetz V, Wiemer DM, Federman SM: Grip and pinch strength: norms for 6- to 19-year-olds. Am J Occup Ther 40:705-11, 1986
81. Kramer M, Heussner P, Herzberg PY, et al: Validation of the grip test and human activity profile for evaluation of physical performance during the intermediate phase after allogeneic hematopoietic stem cell transplantation. Support Care Cancer 21:1121-9, 2013
82. Herzberg PY, Heussner P, Mumm FH, et al: Validation of the human activity profile questionnaire in patients after allogeneic hematopoietic stem cell transplantation. Biol Blood Marrow Transplant 16:1707-17, 2010
83. Akpek G, Zahurak ML, Piantadosi S, et al: Development of a prognostic model for grading chronic graft-versus-host disease. Blood 97:1219-26., 2001
84. Jacobsohn DA, Arora M, Klein JP, et al: Risk factors associated with increased nonrelapse mortality and with poor overall survival in children with chronic graft-versus-host disease. Blood 118:4472-9, 2011
85. Wiklund I: Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. Fundam Clin Pharmacol 18:351-63, 2004
86. 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 1:4, 2003
87. Varni JW, Seid M, Rode CA: The PedsQL: measurement model for the pediatric quality of life inventory. Med Care 37:126-39, 1999
88. Oberg JA, Bender JG, Morris E, et al: Pediatric allo-SCT for malignant and non-malignant diseases: impact on health-related quality of life outcomes. Bone Marrow Transplant 48:787-93, 2013

89. 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 117:4651-7, 2011

90. 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 96:1528-35, 2011

91. Wolff D, Herzberg P, Heussner P, et al: Validation of the NIH criteria of the severity grading of chronic GVHD - results of a prospective German multicenter validation trial. [Abstract] American Society of Hematology 2009; 204. Blood, 2009

92. Inamoto Y, Martin PJ, Chai X, et al: Clinical benefit of response in chronic graft-versus-host disease. Biol Blood Marrow Transplant 18:1517-24, 2012

93. Norman GR, Sloan JA, Wyrwich KW: Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 41:582-92, 2003

94. Norman GR, Sloan JA, Wyrwich KW: Is it simple or simplistic? Med Care 41:599-600, 2003

95. 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

96. Ware JE, Snow KK, Kosinski M, et al: SF-36 Health Survey: a manual and interpretation guide. Boston, The Health Institute, New England Medical Center, 1993

97. Felson DT, Anderson JJ, Boers M, et al: American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 38:727-35, 1995

98. 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 70:404-13, 2011

99. Best WR, Becktel JM, Singleton JW, et al: Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 70:439-44, 1976

100. 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 122:512-30, 2002

101. Anderson JJ, Baron G, van der Heijde D, et al: Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 44:1876-86, 2001

102. Bombardier C, Gladman DD, Urowitz MB, et al: Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum 35:630-40, 1992

103. Liang MH, Socher SA, Larson MG, et al: Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum 32:1107-18, 1989

104. Gladman DD, Ibanez D, Urowitz MB: Systemic lupus erythematosus disease activity index 2000. J Rheumatol 29:288-91, 2002

105. 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 63:3918-30, 2011

106. Luijten KM, Tekstra J, Bijlsma JW, et al: The Systemic Lupus Erythematosus Responder Index (SRI); a new SLE disease activity assessment. Autoimmun Rev 11:326-9, 2012

107. 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 86:447-58, 1993

108. 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 10:541-7, 1992

109. Inamoto Y, Storer BE, Lee SJ, et al: Failure-free survival after second-line systemic treatment of chronic graft-versus-host disease. Blood 121:2340-6, 2013

110. Schultz K, Miklos D, Fowler D: Towards biomarkers for chornic graft versus host disease. Biol Blood Marrow Transplant (in press), 2005

## NIH Consensus Development Project on Criteria for Clinical Trials in Chronic GVHD Steering Committee

## Acknowledgments

## Disclaimer

The opinions expressed here are those of the authors and do not represent the official position of the NIH, FDA or the United States Government.

**Table 1. 2014 changes to the 2005 recommendations**

| 2005 Recommendation | 2014 Recommendation |
|---|---|
| **Organ Measures** | |
| Skin response measured using the body surface area of erythematous rash, moveable sclerosis and non-moveable sclerosis | Skin response measured using the updated NIH Skin Score<br>Detailed collection of type of BSA involvement no longer collected |
| Size of skin ulcers captured | Presence or absence, not size, of skin ulcer captured |
| Eye response measured by change in Schirmer test | Eye response measured by change in NIH Eye Score |
| Mouth response measured by change in the Modified Oral Mucositis Score. Scores range from 0-15 | Remove mucoceles from the Modified Oral Mucositis Score. Scores range from 0-12 |
| Oral chronic GVHD described as “hyperkeratosis” changes | The term “hyperkeratosis” replaced by “lichen-like” changes |
| Patients symptoms of mouth dryness and mouth pain captures on 0-10 scales | No longer recommended. Mouth sensitivity still captured on a 0-10 scale. |
| Change from a 0 to 1 in the NIH Lower GI response measure considered progression | Change from a 0 to 1 in the NIH Lower GI response measure no longer considered progression |
| Liver response measured by change in ALT, bilirubin and alkaline phosphatase | Liver response measured by change in ALT and bilirubin not alkaline phosphatase |
| Lung response measured by change in FEV1 and DLCO after calculation of the Lung Function Score | Lung response measured by change in FEV1 is sufficient for BOS |
| Joints and Fascia not included in response assessment | Use the NIH Joint and Fascia Score and the P-ROM to assess joint response |
| Platelet count and absolute eosinophil count collected to measure hematologic response | Platelet count and absolute eosinophil count collected at baseline only to provide prognostic information |
| All abnormalities captured and attributed to chronic GVHD | All abnormalities captured unless there is another well documented non-chronic GVHD cause, in which case the organ is not evaluable |
| **Ancillary Measures** | |
| Pediatric surveys CHRI and ASK recommended | No longer recommended |
| SF36, FACT-BMT and HAP recommended | SF36 OR FACT-BMT plus HAP are strongly encouraged |
| Two minute walk test recommended | Two minute walk test provides prognostic information, consider assessing at baseline only |
| Grip strength recommended | No longer recommended |
| Karnofsky or Lansky performance status recommended | Karnofsky or Lansky performance status strongly encouraged at baseline only |
| **Response Assignments** | |
| Mixed response category not recognized | Mixed response category recognized and considered progression |
| No recognition of “irreversible” baseline organ damage | Irreversible baseline organ damage may be defined initially and then excluded from the requirements for achieving response |
| No comment on whether responses can be assessed in the setting of additional systemic or organ-directed treatments | Addition of topical or organ-directed treatments for the eyes, mouth, esophagus and genital tract generally make it impossible to assess the response to systemic treatment. |

**Table 2. 2014 Recommended chronic GVHD-specific core measures for assessing responses in chronic GVHD trials**

| Measure | Clinician Assessed | Patient Reported |
|---|---|---|
| Assessments | NIH Skin Score (0-3)<br>NIH Eye Score* (0-3)<br>Modified oral mucositis scale (0-12)<br>Total bilirubin (mg/dL), ALT (U/L)<br>FEV-1 (Liters, % predicted)<br>NIH Joint Score (0-3)<br>Photographic range of motion (4-25) | N/A |
| Symptoms | NIH Lung Symptom Score (0-3)<br>Upper GI Response (0-3)<br>Lower GI Response (0-3)<br>Esophagus Response (0-3) | Lee Symptom Scale [6] (0-100)<br>Skin itching (0-10)<br>Mouth sensitivity (0-10)<br>Chief eye complaint (0-10) |
| Global ratings | None-Mild-moderate-severe [6] (0-3)<br>0-10 severity scale [7] (0-10)<br>7 point change scale [8] (-3 to +3) | None-Mild-moderate-severe[6] (0-3)<br>0-10 severity scale [7] (0-10)<br>7 point change scale [8] (-3 to +3) |

* Components include both signs and symptoms
ALT, alanine transaminase; FEV-1, forced expiratory volume, first second; GI, gastrointestinal;

**Table 3. Strongly encouraged, exploratory, and no longer recommended response measures for general chronic GVHD trials**

| Organ | Strongly Encouraged | Exploratory | No longer recommended for general chronic GVHD studies |
|---|---|---|---|
| Skin | | | Pigmentary changes |
| Eyes | | | Schirmer's test |
| Mouth | | | Mucoceles, patient-reported mouth pain and dryness on a 0-10 scale |
| Upper GI | Weight | | |
| Lower GI | Weight | | |
| Liver | | | Aspartate aminotransferase, alkaline phosphatase |
| Lungs | Corrected DLCO, FVC, TLC, RV | | |
| Hematologic | | | Platelet count, absolute eosinophil count |
| Genitals | | Female and male self-reported question: "Worse genital discomfort" on a 0-10 scale | |
| Ancillary measures | SF-36v.2[9,10,13] (0-100) or FACT-BMT[11] (0-148) in adults HAP [12] (if the SF-36v2 is not captured) (0-94) | PedsQL<br>Clinician and patient-reported severity (0-10) and change (-3 to +3) for organ-specific chronic GVHD manifestations | |

***** No measures for esophagus or joints and fascia

SF-36v2, Medical Outcome Study Short Form 36, version 2; FACT-BMT, Functional Assessment of Cancer Therapy-Bone Marrow Transplantation subscale; HAP, human activities profile; PedsQL, pediatric quality of life

**Table 4. Response determination for chronic GVHD clinical trials based on clinician assessments**

| Organ | Complete Response | Partial Response | Progression | Not evaluable |
|---|---|---|---|---|
| Skin | NIH Skin Score 0 after previous involvement | Decrease in NIH Skin Score by 1 or more points | Increase in NIH Skin Score by 1 or more points | -Skin abnormalities are not due to chronic GVHD<br>-Skin abnormalities are designated irreversible at the start of the trial (not evaluable for improvement)<br>-Skin is still evaluable if topical treatments are added, e.g., topical steroids, topical tacrolimus, moisturizers etc. |
| Eyes | NIH Eye Score 0 after previous involvement | Decrease in NIH Eye Score by 1 or more points | Increase in NIH Eye Score by 1 or more points | -Eye abnormalities are not due to chronic GVHD<br>-Eye abnormalities are designated irreversible at the start of the trial (not evaluable for improvement)<br>-Addition of topical treatments for eye disease, e.g., ophthalmic steroids, cyclosporine or other agents, punctal plugs, special eyewear or contact lenses Does not include autologous serum eyedrops |
| Mouth | Modified Oral Mucositis Score 0 after previous involvement | Decrease in Modified Oral Mucositis Score of 2 or more points | Increase in Modified Oral Mucositis Score of 2 or more points | -Mouth abnormalities are not due to chronic GVHD<br>-Addition of topical treatments for mouth disease, e.g., topical or injected steroids, oral PUVA etc. Does not include artificial saliva, pilocarpine or cimevuline |
| Esophagus | NIH Esophagus Score 0 after previous involvement | Decrease in NIH Esophagus Score by 1 or more points | Increase in NIH Esophagus Score by 1 or more points | -Esophagus abnormalities are not due to chronic GVHD<br>-Esophageal dilation |
| Upper GI | NIH Upper GI Score 0 after previous involvement | Decrease in NIH Upper GI Score by 1 or more points | Increase in NIH Upper GI Score by 1 or more points | -Upper GI abnormalities are not due to chronic GVHD<br>-Addition of topical treatments for upper GI disease, e.g., topical steroids |
| Lower GI | NIH Lower GI Score 0 after previous involvement | Decrease in NIH Lower GI Score by 1 or more points | Increase in NIH Lower GI Score by 1 or more points, except from 0 to 1 | -Lower GI abnormalities are not due to chronic GVHD<br>-Addition of topical treatments for lower GI disease, e.g., topical steroids |

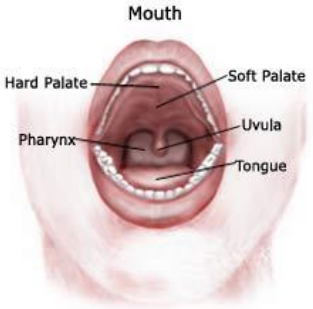
| Organ | Complete Response | Partial Response | Progression | Not evaluable |
|---|---|---|---|---|
| Liver | Normal ALT and Total bilirubin after previous elevation of one or both | -Decrease by 50% if baseline ALT or AST is 3 or more times ULN or total bilirubin more than 3 mg/dL<br>-If baseline ALT or AST is less than 3 times ULN or total bilirubin is 3 mg/dL or less, only CR is possible | -Increase by 2x ULN if baseline ALT or AST is less than 3 times ULN or total bilirubin is more than 3 mg/dL<br>-Increase by 3x ULN if baseline ALT or AST is 3 or more times ULN or total bilirubin is 3 mg/dL or less | -Liver abnormalities are not due to chronic GVHD<br>Liver is still evaluable if ursodeoxycholic acid is added |
| Lungs | -Normal FEV1 after previous involvement<br>-If PFTs not available, NIH Lung Symptom Score 0 after previous involvement | -Increase by 10% absolute value of FEV1<br>-If PFTs not available, decrease in NIH Lung Symptom Score by 1 or more points | -Decrease by 10% absolute value of FEV1<br>-If PFTs not available, increase in NIH Lung Symptom Score by 1 or more points | -Lung abnormalities are clearly not due to chronic GVHD<br>-Lung abnormalities are designated irreversible at the start of the trial (not evaluable for improvement)<br>-Addition of montelukast, which is considered a systemic treatment<br>Does not include addition of azithromycin, inhaled beta agonists or steroids, inhaled tiotropium |
| Joints and Fascia | Both NIH Joint and Fascia Score 0 and P-ROM score 25 after previous involvement by at least one measure | Decrease in NIH Joint and Fascia Score by 1 or more points or increase in P-ROM score by 2 points for shoulder elbow or wrist/finger or 1 or more points for ankle | Increase in NIH Joint and Fascia Score by 1 or more points or decrease in P-ROM score by 2 points for shoulder elbow or wrist/finger or 1 or more points for ankle | -Joint and fascial abnormalities are not due to chronic GVHD<br>-Joint and fascial abnormalities are designated irreversible at the start of the trial (not evaluable for improvement) |
| Global | Clinician overall severity score 0 | Clinician overall severity score decreases by 2 or more points on a 0-10 scale | Clinician overall severity score increases by 2 or more points on a 0-10 scale | |

**FORM A**

**Current Patient Weight:** _____ Today's Date: _____ MR#/Name: _____

## CHRONIC GVHD ACTIVITY ASSESSMENT- CLINICIAN

**Mouth**




| **Mucosal change** | No evidence of cGvHD | | Mild | | Moderate | | Severe | |
|---|---|---|---|---|---|---|---|---|
| **Erythema** | None | 0 | Mild erythema or moderate erythema (<25%) | 1 | Moderate (≥25%) or Severe erythema (<25%) | 2 | Severe erythema (≥25%) | 3 |
| **Lichenoid** | None | 0 | Lichen-like changes (<25%) | 1 | Lichen-like changes (25-50%) | 2 | Lichen-like changes (>50%) | 3 |
| **Ulcers** | None | 0 | | | Ulcers involving (≤20%) | 3 | Severe ulcerations (>20%) | 6 |
| | | | | | | | | |
| | | | | | | | **Total score for all mucosal changes** | |

| **Liver Values** | Total serum bilirubin mg/dL | ULN mg/dL | ALT U/L | ULN U/L |
|---|---|---|---|---|

| | |
|---|---|
| **Gastrointestinal-Upper GI**<br>• Early satiety OR<br>• Anorexia OR<br>• Nausea & Vomiting | *0= no symptoms*<br>*1=mild, occasional symptoms, with little reduction in oral intake during the past week*<br>*2=moderate, intermittent symptoms, with some reduction in oral intake during the past week*<br>*3=more severe or persistent symptoms throughout the day, with marked reduction in oral intake, on almost every day of the past week* |
| **Gastrointestinal-Esophageal**<br>• Dysphagia OR<br>• Odynophagia | *0= no esophageal symptoms*<br>*1=Occasional dysphagia or odynophagia with solid food or pills during the past week*<br>*2=Intermittent dysphagia or odynophagia with solid foods or pills, but not for liquids or soft foods, during the past week*<br>*3=Dysphagia or odynophagia for almost all oral intake, on almost every day of the past week* |
| **Gastrointestinal-Lower GI**<br>• Diarrhea | *0= no loose or liquid stools during the past week*<br>*1= occasional loose or liquid stools, on some days during the past week*<br>*2=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*<br>*3=voluminous diarrhea on almost every day of the past week, **requiring** intervention to prevent or correct volume depletion* |

| **Lungs (Liters and % predicted)**<br>• Bronchiolitis Obliterans | FEV1 | FVC | Single Breath DLCO (adjusted for hemoglobin) | TLC | RV |
|---|---|---|---|---|---|

**Health Care Provider Global Ratings:**

0=none
1= mild
2=moderate
3=severe

**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**

**Over the <<time>> would you say that this patient's cGvHD is**
+3= Very much better
+2= Moderately better
+1= A little better
0= About the same
-1=A little worse
-2=Moderately worse
-3=Very much worse

| **Baseline Values** | Total Distance Walked in 2 Minutes: | Karnofsky or Lansky | Platelet Count K/uL | Total WBC K/uL | Eosinophils % |
|---|---|---|---|---|---|

## CHRONIC GVHD ACTIVITY ASSESSMENT- CLINICIAN (FORM A)

| | SCORE 0 | SCORE 1 | SCORE 2 | SCORE 3 |
|---|---|---|---|---|
| **SKIN**<br><br>**% BSA** ☐ | ☐ No BSA involved | ☐ 1-18% BSA | ☐ 19-50% BSA | ☐ >50% BSA |
| **SKIN FEATURES SCORE:** | ☐ No sclerotic features | | ☐ **S**uperficial sclerotic features "not hidebound" (able to pinch) | **Check all that apply:**<br>☐ Deep sclerotic features<br>☐ "Hidebound" (unable to pinch)<br>☐ Impaired mobility<br>☐ Ulceration |
| ☐ Another documented reason for skin abnormality (specify):_____ | | | | |
| **EYES** | ☐ No symptoms symptoms | ☐ Mild dry eye symptoms not affecting ADL (requirement of lubricant eye drops ≤ 3 x per day) | ☐ Moderate dry eye symptoms partially affecting ADL (requiring lubricant eye drops > 3 x per day or punctal plugs), **WITHOUT** new vision impairment due to KCS | ☐ Severe dry eye symptoms significantly affecting ADL (special eyeware to relieve pain) **OR** unable to work because of ocular symptoms **OR** loss of vision due to KCS |
| ☐ Another documented reason for eye abnormality (specify):_____ | | | | |
| **LUNGS** | ☐ 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 $0_2$) |
| ☐ Another documented reason for lung abnormality (specify):_____ | | | | |
| **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.) |
| ☐ Another documented reason for joint/fascia abnormality (specify):_____ | | | | |

## CHRONIC GVHD ACTIVITY ASSESSMENT- CLINICIAN

**Shoulder**


❑ Not done

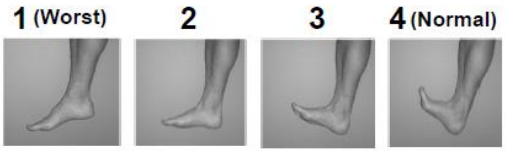
**Elbow**


❑ Not done

**Wrist/finger**


❑ Not done

**Ankle**




❑ Not done

**FORM B**

Today's Date: _____ MR#/Name: _____

# CHRONIC GVHD ACTIVITY ASSESSMENT-PATIENT SELF REPORT

| **Symptoms**<br>**Please rate how severe the following symptoms have been in the last seven days. Please fill in the circle below from 0 (symptom has not been present) to 10 (the symptom was as bad as you can imagine it could be) for each item.** | **Not Present**<br>0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | **As Bad As You Can Imagine**<br>10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Your **skin itching** at its WORST? | O | O | O | O | O | O | O | O | O | O | O |
| Your **mouth sensitivity** at its WORST? | O | O | O | O | O | O | O | O | O | O | O |
| Your **genital discomfort** at its WORST? (Women – vagina, vulva, or labia) (Men – penis) | O | O | O | O | O | O | O | O | O | O | O |

| | | |
|---|---|---|
| Eyes | What is your main complaint with regard to your eyes? | |
| | Please rate how severe this symptom is, from 0 (not at all severe) to 10 (most severe): | 0 1 2 3 4 5 6 7 8 9 10 |

**Patient Global Ratings:**

**1. Overall, do you think that your chronic graft versus host disease is mild, moderate or severe?**
1= mild
2=moderate
3=severe

**2. Please circle the number indicating how severe your chronic graft versus host disease symptoms are, where 0 is cGvHD symptoms that are not at all severe and 10 is the most severe chronic GvHD symptoms possible.**

0 1 2 3 4 5 6 7 8 9 10

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

**3. Compared to a month ago, overall would you say that your cGvHD symptoms are:**

+3= Very much better
+2= Moderately better
+1=A little better
0= About the same
-1=A little worse
-2=Moderately worse
-3=Very much worse