Published in final edited form as:
*Arthritis Rheumatol*. 2014 April ; 66(4): 1044–1052. doi:10.1002/art.38293.

# Joint and fascia manifestations in chronic graft-versus-host disease and their assessment

**Yoshihiro Inamoto, MD**[1], **Joseph Pidala, MD**[2], **Xiaoyu Chai, MS**[1], **Brenda F. Kurland, PhD**[1,3], **Daniel Weisdorf, MD**[4], **Mary E.D. Flowers, MD**[1], **Jeanne Palmer, MD**[5], **Sally Arai, MD**[6], **David Jacobsohn, MD**[7], **Corey Cutler, MD**[8], **Madan Jagasia, MD**[9], **Jenna D. Goldberg, MD**[10], **Paul J. Martin, MD**[1], **Steven Z. Pavletic, MD**[11], **Georgia B. Vogelsang, MD**[12], **Stephanie J. Lee, MD**[1], and **Paul A. Carpenter, MD**[1] **on behalf of the Chronic GVHD Consortium**

[1]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

[2]Blood and Marrow Transplantation, Moffitt Cancer Center, Tampa, FL

[3]Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA

[4]Blood and Marrow Transplant Program, University of Minnesota, Minneapolis, MN

[5]Division of Hematology/Oncology, Medical College of Wisconsin, Milwaukee, WI

[6]Division of Blood and Marrow Transplantation, Stanford University Medical Center, Stanford, CA

[7]Division of Blood and Marrow Transplantation, Children's National Medical Center, Center for Cancer and Blood Disorders, Washington, DC

[8]Hematologic Malignancies, Dana-Farber Cancer Institute, Boston, MA

[9]Hematology and Stem Cell Transplant Program, Vanderbilt University Medical Center, Nashville, TN

[10]Adult Bone Marrow Transplantation Service, Memorial Sloan-Kettering Cancer Center, New York, NY

[11]National Cancer Institute, Bethesda, MD

[12]Oncology, Johns Hopkins Hospital, Baltimore, MD, USA

## Abstract

**Address Correspondence to** Paul A. Carpenter, MD, Fred Hutchinson Cancer Research Center, D5-290, PO Box 19024, Seattle, WA 98109-1024; Telephone: (206)-667-5191; Fax: (206)-667-1034; pcarpent@fhcrc.org..

**Conflict of interest:** The authors declare no competing financial interests related to this study.

**AUTHOR CONTRIBUTIONS** All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Lee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Yoshihiro Inamoto, Brenda Kurland, Stephanie Lee, Paul Carpenter.

**Acquisition of data.** Joseph Pidala, Mary Flowers, Jeanne Palmer, David Jacobsohn, Corey Cutler, Madan Jagasia, Stephanie Lee, Paul Carpenter.

**Analysis and interpretation of data.** Yoshihiro Inamoto, Joseph Pidala, Xiaoyu Chai, Brenda Kurland, Daniel Weisdorf, Sally Arai, Jenna Goldberg, Paul Martin, Steven Pavletic, Georgia Vogelsang, Stephanie Lee, Paul Carpenter.

**Objective—**Joint and fascia manifestations in patients with chronic graft-versus-host disease (GVHD) after allogeneic hematopoietic cell transplantation need to be assessed reliably, simply and in a clinically meaningful way.

**Methods—**In a prospective, multicenter, longitudinal, observational cohort of patients with chronic GVHD (n=567), we evaluated 3 scales proposed for assessing joint status: National Institutes of Health (NIH) joint/fascia scale, Hopkins fascia scale and the Photographic Range of Motion (P-ROM) scale. Ten other scales were also tested for assessing symptoms, quality of life and physical functions.

**Results—**Joint and fascia manifestations were present at study enrollment in 164 (29%) patients. Limited range of motion was most frequent at wrists or fingers. Among the 3 joint scales, changes in the NIH scale correlated with both clinician and patient-perceived improvement of joint and fascia manifestations with higher sensitivity than the Hopkins fascia scale. Changes in all 3 scales correlated with clinician and patient-perceived worsening but the P-ROM scale was the most sensitive in this regard. Onset of joint and fascia manifestations was not associated with subsequent mortality.

**Conclusion—**Joint and fascia manifestations are common and should be assessed carefully in patients with chronic GVHD. Our results support the use of the NIH joint/fascia scale and P-ROM scale to assess joint and fascia manifestations. The NIH scale better captures improvement, while the P-ROM scale better captures worsening. The utility of these scales could also be tested in the rheumatic diseases.

Allogeneic hematopoietic cell transplantation is a curative treatment for many hematologic diseases.(1) Chronic graft-versus-host disease (GVHD) occurs in approximately half of the transplant survivors and is the leading cause of late morbidity that compromises quality of life (QOL) and function.(2-4) Chronic GVHD is thought to occur because the donor's immune system recognizes recipient tissues, causing inflammation and fibrosis. Although joint/fascia manifestations have been considered to be infrequent in patients with chronic GVHD, studies investigating this complication have been limited. Reported joint/fascia manifestations include joint stiffness, edema, restricted range of motion (ROM), arthralgia and rarely arthritis or synovitis.(5) Joint/fascia manifestations may be clinically detectable when inflammation and fibrosis arise in deep tissues (deep sclerosis / fasciitis) or skin overlying joints (superficial sclerosis), and the former may occur with or without superficial sclerosis.(5) Isolated fasciitis is frequently recognizable by restricted ROM or joint contractures. It is usually accompanied by stiffness or edema of the extremities, while the overlying skin remains freely mobile.(5) For example, inability to assume a "Buddha prayer" posture with full bilateral wrist extension indicates limited wrist extension due to tightening of flexor tendons. Sometimes superficial sclerosis is confluent with deep sclerosis or fasciitis, in which case, the skin may be hidebound and underlying tissue has a wooden texture.

Joint/fascia manifestations in patients with chronic GVHD need to be assessed reliably, simply and in a clinically meaningful way. Severity of joint/fascia manifestations and response to therapy require documentation both in clinical trials and clinical practice to guide therapy. Recognizing the lack of validated joint assessment scales, the 2005 National

Institutes of Health (NIH) Consensus conferees and other investigators proposed several measurement scales.(5-13) In order to determine the optimal approach for capturing changes in joint/fascia manifestations in patients with chronic GVHD, we evaluated 3 joint assessment scales and 10 other scales that assess symptoms, QOL and physical functions. We also examined longitudinal joint responses according to the validated scales and associations of joint/fascia manifestations with subsequent mortality.

## PATIENTS AND METHODS

### Study cohort

Patients who were at least two years of age, with systemically treated chronic GVHD within 3 years after transplantation, were eligible for a prospective, multicenter, longitudinal, observational study of the Chronic GVHD Consortium.(14) Patients with recurrent disease or anticipated survival less than 6 months were not eligible. Diagnosis of chronic GVHD was made according to the NIH consensus criteria.(5) Incident (enrollment <3 months after chronic GVHD diagnosis) and prevalent (enrollment ≥3 months after chronic GVHD diagnosis but within 3 years after transplantation) cases were included. At enrollment and every 6 months thereafter, clinicians and patients reported standardized information on chronic GVHD organ involvement and manifestations. Incident cases had an additional assessment at 3 months after enrollment. Patients were treated according to institutional practice in compliance with the NIH chronic GVHD consensus guidelines.(5) The study protocol was approved by the Institutional Review Board of each participating center, and all participants or their guardians gave written informed consent in accordance with the Declaration of Helsinki.

### Assessment scales

A total of 13 assessment scales were evaluated in this study (Table 1). The NIH joint/fascia scale uses a 0-3 point scale to score a composite of tightness, ROM and activities of daily living (ADL). The Hopkins fascia scale uses a 0-3 scale but scores only tightness. The Photographic Range of Motion (P-ROM) scale is a series of images that captures ROM separately for shoulders, elbows, wrists/fingers, and ankles (Supplementary Figure S1).(7) Lower scores indicate more limited ROM. The P-ROM total score is the sum of scores in all 4 joints for a maximum of 25 points. P-ROM data were collected among 502 patients who were enrolled after November 2008. The Lee symptom scale is a 30-item self-administered patient questionnaire specific to symptoms of chronic GVHD.(8) The muscle/joint subscale from the Lee overall symptom scale was also evaluated in this study. Patients reported their overall chronic GVHD symptoms on a 10-point scale for peak severity during the past week.(9) The Functional Assessment of Cancer Therapy-General (FACT-G) and the Short Form 36 (SF36) were evaluated for QOL assessment.(10, 11) The Human Activities Profile (HAP) is a 94-item self-reported assessment of energy expenditure or physical fitness.(12) The walk test is for measuring physical performance by the total distance in feet walked in 2 minutes.(15) The grip strength test is for measuring physical performance by a hydraulic dynamometer.(16) Average pounds of pressure from 3 measurements in the dominant hand are used for analysis.

### Statistical analysis

Joint/fascia manifestations were defined as a NIH joint/fascia score ≥1 at any study visit. At follow-up visits every 3 to 6 months, as an anchor of response, both clinician and patient rated separately their perception of change in joint/fascia manifestations on an 8-point scale that was collapsed for analysis into: *improved* ("[1] completely gone", "[2] very much better", "[3] moderately better"), *stable* ("[4] a little better", "[5] about the same", "[6] a little worse"), *or worse* ("[7] moderately worse", "[8] very much worse"). Longitudinal change scores for scales were calculated by subtracting previous values from current values. To account for within patient correlation, multivariable linear mixed models with random patient effect were used to evaluate correlations between changes in each scale and clinician or patient-perceived changes in joint status (*improved* vs. *stable* or *worse* vs. *stable*). The analysis included all paired visits when joint/fascia manifestations were documented in the previous or current visit. Linear mixed models were chosen since the models were little affected by missing data.(17, 18) All models were adjusted by covariates associated with longitudinal changes in measures in univariate analysis at $P$ values ≤0.01. In comparing performance among the different scales, the estimated differences in measures according to clinician- or patient-perceived improvement or worsening (vs. stability) were standardized by clinically meaningful change of the scale. This standardization is important because each scale has a different increment and potential range. As described by the NIH Consensus,(13) clinically meaningful changes were defined according to the original design of the scale or half of the standard deviation of baseline values (Table 1).

Cox regression models were used to examine correlations between onset of joint/fascia manifestations and subsequent overall and nonrelapse mortality, treating onset of joint/fascia manifestations as a time-dependent covariate. The models were adjusted for study site, case type, months from transplantation to enrollment, platelet count, serum total bilirubin, Karnofsky score, prednisone dose, patient age at transplantation, HLA and donor type, donor-recipient gender combination, conditioning intensity, history of grades II-IV acute GVHD and classic or overlap subcategory. These covariates were chosen to control for known chronic GVHD mortality risk factors and potential outcome differences among study sites.(19-22)

Proportions of joint response across time after visits with newly developed joint/fascia manifestations were graphically plotted. Newly developed joint/fascia manifestations were defined as a NIH joint/fascia score ≥1 at enrollment for incident cases, and as the first onset of NIH joint/fascia score ≥1 without previous joint/fascia manifestations for prevalent cases. Statistical analyses were performed using SAS/STAT software, version 9.3 (SAS Institute, Inc., Cary, NC) and R version 2.15.2 (R Foundation for Statistical Computing, Vienna, Austria).

## RESULTS

### Patient characteristics and the presence of joint/fascia manifestations

A total of 567 participants were enrolled through December 2011. The median follow-up time of survivors was 23.6 months (interquartile range [IQR], 13.3-34.0 months) after

enrollment. Table 2 shows characteristics of the 567 patients at the time of enrollment (baseline). Joint/fascia manifestations, as defined by a NIH joint/fascia score ≥1, were present at study enrollment in 164 (29%) patients. Joint/fascia manifestations at enrollment were associated with longer duration from transplantation to enrollment, prevalent cases, and the use of high-dose total body irradiation conditioning. Other characteristics were similar between the two groups. Chronic GVHD characteristics were also compared between patients with and without joint/fascia manifestations at enrollment (Table 3). In this context, joint/fascia manifestations were associated with more frequent skin involvement and skin sclerosis, less frequent mouth and liver involvement, higher NIH global severity score, higher symptom burden and lower QOL as measured by the FACT-G and SF36-Physical Component Score. SF36-Mental Component Score, maximum and adjusted HAP scores, walk test and grip test results were similar between the two groups. Walk test results did not differ between patients with limited ROM in ankles at enrollment and those without joint/fascia manifestations (median 466 feet [IQR 400-536] vs. 500 feet [IQR 410-575], $P = 0.08$). Grip test results were lower among patients with limited ROM in wrists/fingers at enrollment than those without joint/fascia manifestations (median 51 lb [IQR 42.7-75.3] vs. 62.3 lb [IQR 49.7-81], $P = 0.02$).

Among the 164 patients with joint/fascia manifestations at enrollment, 107 (65%) had mild joint/fascia manifestations, 51 (31%) had moderate manifestations and 6 (4%) had severe manifestations according to the NIH joint/fascia score. Among 98 patients with joint/fascia manifestations and available P-ROM data at enrollment, limitations in ROM were present in wrists/fingers (64%), ankles (47%), shoulders (35%) and elbows (30%)(Figure 1). Limitations in ROM were most frequently mild in all joints according to the P-ROM score (i.e., score 6 or 5 for shoulders, elbows and wrists/fingers, and score 3 for ankles), and limitations were present in multiple joints for 72% of the patients with limited ROM in at least one joint. The median and mean of the P-ROM total score at enrollment were 25 (IQR, 24-25) and 23.9 (standard deviation, 2.1), respectively.

### Difference in longitudinal changes in measurement scores according to perceived changes at follow-up visits

Changes in joint status were examined for 652 paired visits when joint/fascia manifestations were documented in the previous or current visit. In later visits, changes in joint status were rated by clinicians as improved in 44%, stable in 51% and worse in 5%, and by patients as 45%, 44% and 11%, respectively. Agreement between clinicians and patients was moderate (weighted kappa = 0.32). Patients tended to report more improvement and worsening than clinicians.

Estimated differences in longitudinal changes in measures between improvement and stability or between worsening and stability for the 3 joint/fascia scales are shown in Figure 2A. The “estimated difference” in linear mixed models indicates the average difference in scores for the group of visits associated with perceived improvement or perceived worsening as compared to the group of visits associated with perceived stability (see Supplementary Figure S2 for details). For example, the NIH joint/fascia score improved by an estimated

average of 0.41 point (95% confidence interval [CI] 0.28-0.55, $P < 0.001$) when clinicians perceived improvement vs. stability (Figure 2A left).

Among the 3 joint/fascia scales, changes in the NIH joint/fascia score and Hopkins fascia score correlated with both clinician and patient-perceived joint improvement (Figure 2A left), whereas changes in the P-ROM total score correlated with clinician-perceived improvement but not with patient-perceived improvement. By clinician perception, estimated differences between improvement and stability were larger for the NIH score than for the Hopkins score. Therefore, the NIH joint/fascia scale is more sensitive to clinician-perceived improvement than the Hopkins fascia scale. For patient perception, estimated differences were similar between the NIH score and Hopkins score. In comparing worsening vs. stability, changes in all of the 3 joint/fascia scales correlated with both clinician and patient-perceived joint worsening (Figure 2A right). Among the 3 scales, estimated differences between worsening and stability were significantly larger for the P-ROM total scale than for the other 2 scales by both clinician and patient perception. Therefore, the P-ROM scale is most sensitive to worsening. The NIH joint/fascia score might have had an advantage in demonstrating change since this score was used to select visit pairs, but results were similar even if the P-ROM score was used to select visit pairs for analysis (data not shown).

Estimated standardized differences in scores for other scales are shown in Figure 2B. Changes only in the SF36-PCS correlated with both clinician and patient-perceived joint improvement (Figure 2B left). In comparing worsening vs. stability, changes in all 3 symptom scores and FACT-G scores correlated with both clinician and patient-perceived joint worsening (Figure 2B right). Changes in the HAP scores correlated with clinician-perceived worsening but not with other perceived changes. There were no statistical associations of changes in walk test or grip strength test results with clinician- or patient-perceived changes in the joints, and the results were similar even if the analysis was limited to only patients with limited ROM in ankles or wrists.

## Longitudinal response assessment according to the NIH joint/fascia scale and PROM scale

Seventy-seven percent (108/140) of patients in our study cohort with new joint/fascia manifestations had subsequent visits at 3 or 6 months. Joint response according to the NIH joint/fascia and P-ROM scales is shown in Figure 3. Analysis beyond 6 months was not interpretable because more than half of data were missing. Among incident cases (Figure 3A and B), there was little difference between 3 and 6 months in the proportions of patients categorized as having joint improvement, stability and worsening, according to both scales, suggesting that the changes were evident by 3 months after onset of joint/fascia manifestations. Among incident cases (Figure 3A vs. B), improvement was approximately 10% lower using the P-ROM scale compared to the NIH scale, while worsening was 10-15% higher using the P-ROM scale compared to the NIH scale. This trend was more obvious among prevalent cases than among incidence cases. Compared with incident cases, improvement was less frequent among prevalent cases, while worsening was more frequent among prevalent cases (Figure 3A vs. C and B vs. D).

### Association of joint/fascia manifestations with survival outcomes

In multivariable time-dependent Cox models, joint/fascia manifestations at any time (NIH joint/fascia score ≥1) were not associated with subsequent overall mortality or nonrelapse mortality (results not shown). Results were similar when only moderate or severe joint/fascia manifestations (NIH joint/fascia score ≥2) were considered. The number of patients with severe manifestations was not sufficient for a separate analysis of this group.

## DISCUSSION

Our results showed a 29% incidence of joint/fascia manifestations in patients with chronic GVHD. Although the cohort did not include all consecutive patients at each participating center and therefore a selection bias may be present, we believe that our results emphasize the importance of careful examination of the joints and fasciae in this population. Based on our data, it is particularly important to provide education about potential joint/fascia manifestations among patients who are more than 1 year after transplantation, those who received high-dose total body irradiation conditioning, or those who had skin involvement or sclerosis with GVHD.(23, 24)

The NIH joint/fascia score was originally intended to evaluate the severity of GVHD manifestations in joints and fasciae for baseline or cross-sectional use,(5) but our results suggest that longitudinal changes in the NIH joint/fascia score between visits could be used for evaluating response. Recent studies showed similar utility of longitudinal changes in the NIH organ score for measuring response in the skin and eyes.(25, 26) Changes in the Hopkins fascia score also correlated with improvement and worsening from clinician and patient perspectives, but estimated differences were smaller for the Hopkins fascia score than those for the NIH joint/fascia score, indicating less sensitivity of the Hopkins fascia scale which may be explained by differences in what it captures. The NIH joint/fascia score incorporates all three domains of tightness, ROM and ADL, whereas the Hopkins fascia score addresses only tightness. Thus we recommend that the Hopkins fascia score can be omitted if the NIH score is collected.

The merit of the P-ROM scale is its objectivity and simplicity.(7) Active-assisted ROM was recommended as a useful objective measure of joint response by the NIH Consensus, but the main limitation of this assessment has been the need for an adequately trained professional who can conduct ROM measurements in a standardized and reproducible fashion.(13) Therefore the P-ROM scale was developed as an alternative for clinical use since any provider, including a family physician, can complete the assessment in 1-2 minutes. Although the P-ROM scale was the most sensitive to perceived joint worsening among all scales, it was insensitive to patient-perceived joint improvement, perhaps because the P-ROM scale does not consider tightness or ADL. We often observe patients who report improvement in tightness before we or they observe improvement in ROM, which tends to occur more slowly. Such subtle changes may be more readily apparent to patients than to clinicians. One consideration for the future would be to increase the P-ROM sensitivity by incorporating a tightness component in this scale.

Changes in symptom scales did not correlate with clinician-perceived improvement. Symptom information must be derived from patients, and patients' perceptions are often discordant with clinicians' assessments. In this context, the Lee muscle/joint symptom subscale is useful to capture changes in joint-specific symptoms. Similarly, either of the Lee overall symptom scale or 10-point global rating scale is useful for capturing changes in overall symptoms.

The FACT-G was sensitive to worsening but not to improvement, while the converse was true of the SF36-PCS, suggesting that both scales were not perfectly sufficient to capture changes in QOL associated with joint response. We did not observe a correlation of changes in activity or physical function scales with joint response. These scales may either lack sufficient sensitivity or relevance for being able to detect changes in joint status. Non-articular manifestations of GVHD may have more impact on these measures.

The onset of joint/fascia manifestations was not associated with subsequent mortality outcomes, supporting our understanding that disability and morbidity are more important than mortality in these patients. This result is consistent with another study that showed similar transplant outcomes between chronic GVHD patients with and without sclerotic manifestations except for prolonged duration of immunosuppressive treatment.(24)

This study has some limitations. First, the study population is comprised mostly of adults who received mobilized blood cell grafts. The results may not apply to children or those who received transplantation from other stem cell sources. Second, the scales used herein may not reflect symptoms associated with arthralgia or arthritis. Arthralgia is sometimes observed but is often difficult to document and not captured. In contrast, arthritis with destruction occurs rarely but true incidence data are lacking. Future studies should elucidate the frequency, presentation and significance of these manifestations. Lastly, we were unable to evaluate treatment effect for joint/fascia manifestations since immunosuppressive or physical therapies were not mandated in this observational study. Future prospective interventional studies could address this question using the validated scales.

We report the first attempt to validate scales for assessing joint/fascia manifestations in patients with chronic GVHD. Our results support the use of the NIH joint/fascia scale and Photographic Range of Motion (P-ROM) scale. The NIH scale better captures improvement, while the P-ROM scale better captures worsening. Our longitudinal observation clarified that joint response was evident by 3 month after onset of joint/fascia manifestations, and that significant proportions of patients experienced worsening in ROM within 6 months if joint/fascia manifestations developed later than 3 months after diagnosis of chronic GVHD. The utility of these scales could also be tested in the rheumatic diseases.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Drs Anne Stevens and Mark Wener for reviewing and commenting on the manuscript.

**Financial disclosure:** This work was supported by grants CA118953, CA163438 and CA047904 from the National Institutes of Health (NIH). The Chronic GVHD Consortium (U54 CA163438) is a part of the NIH Rare Diseases Clinical Research Network (RDCRN), supported through collaboration between the NIH Office of Rare Diseases Research (ORDR) at the National Center for Advancing Translational Science (NCATS), the National Cancer Institute, and the Fred Hutchinson Cancer Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES

1. Thomas E, Storb R, Clift RA, Fefer A, Johnson FL, Neiman PE, et al. Bone-marrow transplantation (first of two parts). N Engl J Med. 1975; 292:832–43. [PubMed: 234595]
2. Lee SJ, Vogelsang G, Flowers ME. Chronic graft-versus-host disease. Biol Blood Marrow Transplant. 2003; 9:215–33. [PubMed: 12720215]
3. van den Bergh V, Tricot G, Fonteyn G, Dom R, Bulcke J. Diffuse fasciitis after bone marrow transplantation. Am J Med. 1987; 83:139–43. [PubMed: 3300320]
4. Janin A, Socie G, Devergie A, Aractingi S, Esperou H, Verola O, et al. Fasciitis in chronic graft-versus-host disease. A clinicopathologic study of 14 cases. Ann Intern Med. 1994; 120:993–8. [PubMed: 8185148]
5. Filipovich AH, Weisdorf D, Pavletic S, Socie G, Wingard JR, Lee SJ, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. Biol Blood Marrow Transplant. 2005; 11:945–56. [PubMed: 16338616]
6. Jacobsohn DA, Chen AR, Zahurak M, Piantadosi S, Anders V, Bolanos-Meade J, et al. Phase II study of pentostatin in patients with corticosteroid-refractory chronic graft-versus-host disease. J Clin Oncol. 2007; 25:4255–61. [PubMed: 17878478]
7. Carpenter PA. How I conduct a comprehensive chronic graft-versus-host disease assessment. Blood. 2011; 118:2679–87. [PubMed: 21719600]
8. 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; 8:444–52. [PubMed: 12234170]
9. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. Cancer. 2000; 89:1634–46. [PubMed: 11013380]
10. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol. 1993; 11:570–9. [PubMed: 8445433]
11. McHorney CA, Ware JE Jr. Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993; 31:247–63. [PubMed: 8450681]
12. Herzberg PY, Heussner P, Mumm FH, Horak M, Hilgendorf I, von Harsdorf S, et al. Validation of the human activity profile questionnaire in patients after allogeneic hematopoietic stem cell transplantation. Biol Blood Marrow Transplant. 2010; 16:1707–17. [PubMed: 20541028]
13. Pavletic SZ, Martin P, Lee SJ, Mitchell S, Jacobsohn D, Cowen EW, 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; 12:252–66. [PubMed: 16503494]
14. The Chronic GVHD Consortium. Rationale and Design of the Chronic GVHD Cohort Study: Improving Outcomes Assessment in Chronic GVHD. Biol Blood Marrow Transplant. 2011; 17:1114–20. [PubMed: 21664473]
15. Waters RL, Lunsford BR, Perry J, Byrd R. Energy-speed relationship of walking: standard tables. Journal of orthopaedic research : official publication of the Orthopaedic Research Society. 1988; 6:215–22. [PubMed: 3343627]
16. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. The Journal of hand surgery. 1984; 9:222–6. [PubMed: 6715829]

17. Fitzmaurice, N.; Laird, N.; Ware, J. Applied Longitudinal Analysis. John Wiley & Sons, Inc.; Hoboken, NJ: 2004.

18. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? Stat Med. 2009; 28:221–39. [PubMed: 19012297]

19. Arora M, Klein JP, Weisdorf DJ, Hassebroek A, Flowers ME, Cutler CS, et al. Chronic GVHD risk score: a Center for International Blood and Marrow Transplant Research analysis. Blood. 2011; 117:6714–20. [PubMed: 21493797]

20. Sullivan KM, Witherspoon RP, Storb R, Weiden P, Flournoy N, Dahlberg S, 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; 72:546–54. [PubMed: 3042041]

21. Wingard JR, Piantadosi S, Vogelsang GB, Farmer ER, Jabs DA, Levin LS, et al. Predictors of death from chronic graft-versus-host disease after bone marrow transplantation. Blood. 1989; 74:1428–35. [PubMed: 2670000]

22. Vigorito AC, Campregher PV, Storer BE, Carpenter PA, Moravec CK, Kiem HP, et al. Evaluation of NIH consensus criteria for classification of late acute and chronic GVHD. Blood. 2009; 114:702–8. [PubMed: 19470693]

23. Martires KJ, Baird K, Steinberg SM, Grkovic L, Joe GO, Williams KM, et al. Sclerotic-type chronic GVHD of the skin: clinical risk factors, laboratory markers, and burden of disease. Blood. 2011; 118:4250–7. [PubMed: 21791415]

24. Inamoto Y, Storer BE, Petersdorf EW, Nelson JL, Lee SJ, Carpenter PA, et al. Incidence, risk factors and outcomes of sclerosis in patients with chronic graft-versus-host disease. Blood. 2013; 121:5098–103. [PubMed: 23547053]

25. Inamoto Y, Chai X, Kurland BF, Cutler C, Flowers ME, Palmer JM, et al. Validation of measurement scales in ocular graft-versus-host disease. Ophthalmology. 2012; 119:487–93. [PubMed: 22153706]

26. Jacobsohn DA, Kurland BF, Pidala J, Inamoto Y, Chai X, Palmer JM, et al. Correlation between NIH composite skin score, patient reported skin score and outcome: results from the Chronic GVHD Consortium. Blood. 2012; 120:2545–52. [PubMed: 22773386]

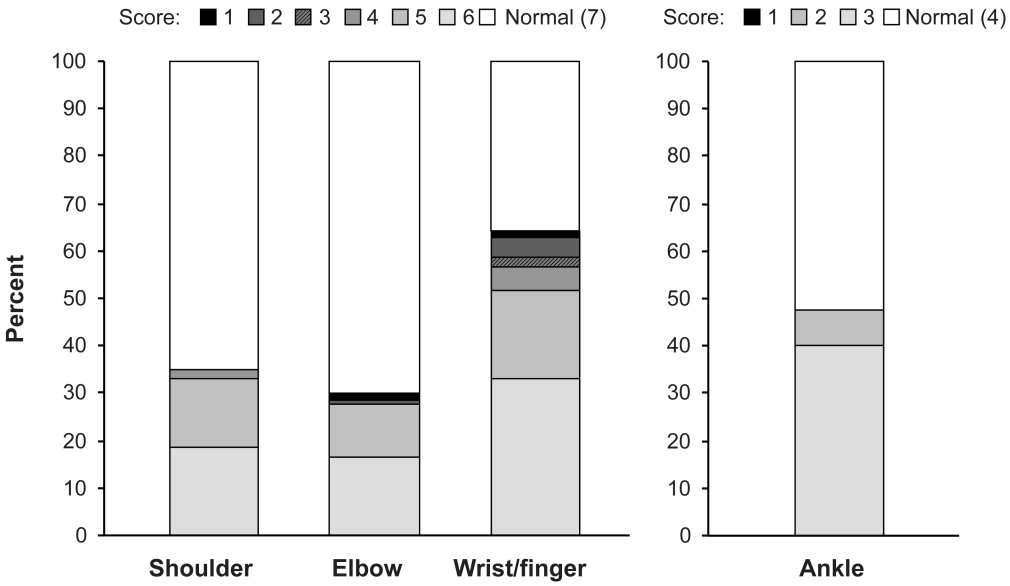


**Figure 1.**

Sites and distribution of P-ROM scores at baseline among 98 patients with joint/fascia manifestations by NIH joint/fascia score ≥1 and available P-ROM. (Full ROM = score 7 in shoulders, elbows, wrists/fingers and score 4 in ankles).

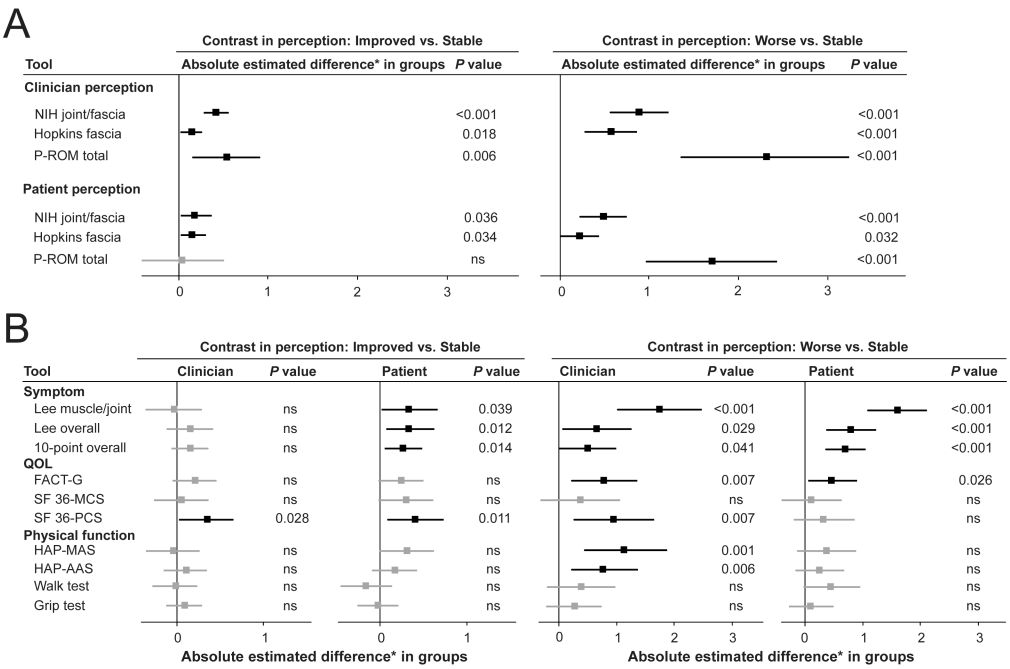


**Figure 2.**

Estimated differences in scores and 95% confidence intervals according to clinician or patient-perceived change in joint/fascia manifestations. Black color indicates statistically significant correlation and grey color indicates statistically non-significant (ns) correlation. All models were adjusted by case type (incident vs. prevalent), which was the sole covariate associated with longitudinal changes in measures in univariate analysis. (**A**) Joint/fascia scales. (**B**) Symptom scales, quality of life scales and physical function scales. NIH = National Institutes of Health; P-ROM = photographic range of motion; QOL = quality of life; FACT-G = Functional Assessment of Cancer Therapy-General; SF36 = Short Form 36; MCS = mental component score; PCS = physical component score; HAP = Human Activities Profile; MAS = maximum activity score; AAS = adjusted activity score. *Estimated differences are standardized by the clinically meaningful change of the scale.

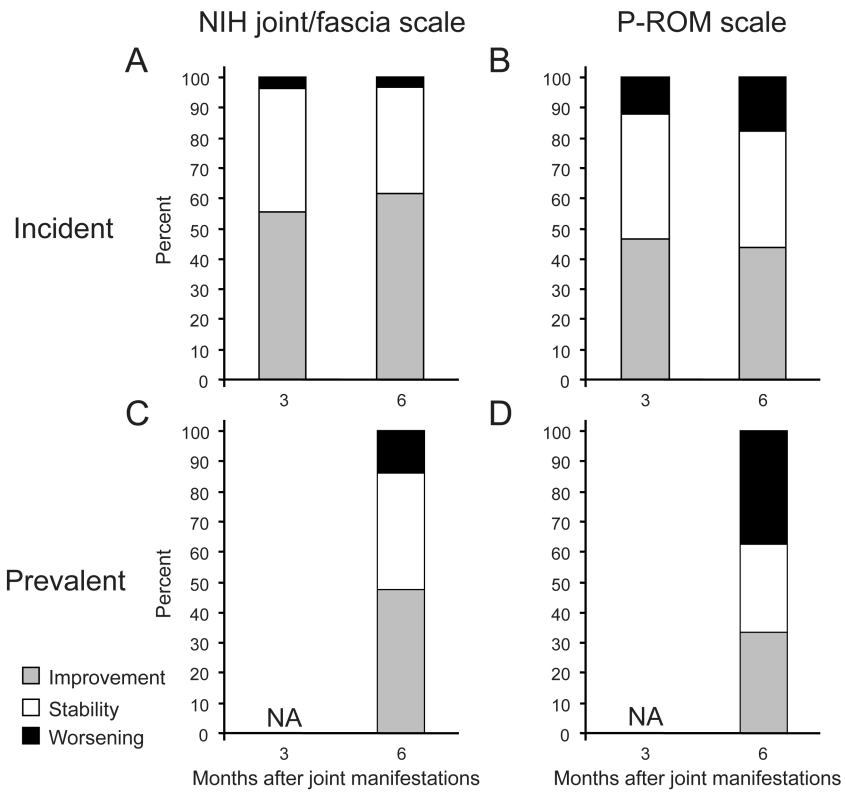


**Figure 3.**

Longitudinal response assessment according to the NIH joint/fascia scale and P-ROM total scale. Proportions of joint response across time after newly developed joint/fascia manifestations are shown in incident cases (**A** and **B**) and prevalent cases (**C** and **D**). Response was not assessed (NA) at 3 months in prevalent cases.

## Table 1

Assessment scales evaluated in this study*

| Assessment scale | Score | No. of items | Component | Baseline SD | Clinically meaningful change |
|---|---|---|---|---|---|
| NIH joint/fascia scale (range, 0-3) | 0: No symptoms<br>1: Mild tightness of arms or legs, normal or mild decreased ROM AND not affecting ADL<br>2: Tightness of arms or legs OR joint contractures, erythema thought due to fasciitis, moderate decrease ROM AND mild to moderate limitation of ADL<br>3: Contracture WITH significant decrease of ROM AND significant limitation of ADL (unable to tie shoes, button shirts, dress self etc.) | 1 | Tightness<br>ROM<br>ADL | NA | 1 point† |
| Hopkins fascia scale (range, 0-3) | 0: Normal<br>1: Tight with normal areas<br>2: Tight<br>3: Tight, unable to move | 1 | Tightness | NA | 1 point† |
| P-ROM scale (range, 4-25) | The summary of the 7-point wrist, shoulder, elbow scales plus the 4-point ankle scale (see Figure S1). | 4 | ROM | 2.1 | 1 point‡ |
| Lee muscle/joint subscale (range, 0-16) | Summary of the following 4 items. Each item is rated at 0: not at all, 1: slightly, 2: moderately, 3: quite a bit, or 4: extremely.<br>- Joint and muscle aches<br>- Limited joint movement<br>- Muscle cramps<br>- Weak muscles | 4 | Symptom | 4.0 | 2 points‡ |
| Lee overall symptom scale (range, 0-100) | 30-item self-administered patient questionnaire specific to symptoms of chronic GVHD. | 30 | Symptom | 13.0 | 6.5 points‡ |
| 10-point global rating (range, 0-10) | Chronic GVHD symptoms overall in the last week. Rated from 0 (not present) to 10 (as bad as you can imagine). | 1 | Symptom | NA | 2 points† |
| FACT-G | 27-item self-report questionnaire, which was validated for measuring response of chronic GVHD. | 27 | QOL | 16.2 | 8.1 points‡ |
| SF36-MCS | Mental component score from 36-item self-report questionnaire assessing health and functioning. | 36 | QOL | 10.9 | 5.5 points‡ |
| SF36-PCS | Physical component score from 36-item self-report questionnaire assessing health and functioning. | 36 | QOL | 9.8 | 4.9 points‡ |
| HAP-MAS | Maximum activity score from 94-item self-reported assessment of energy | 94 | ADL | 12.7 | 6.4 points‡ |

| Assessment scale | Score | No. of items | Component | Baseline SD | Clinically meaningful change |
|---|---|---|---|---|---|
| | expenditure or physical fitness. | | | | |
| HAP-AAS | Adjusted activity score from 94-item self-reported assessment of energy expenditure or physical fitness. | 94 | ADL | 17.3 | 8.7 points‡ |
| Walk test | Total distance walked in 2 minutes. | 1 | Physical function | 128.4 | 64 feet‡ |
| Grip test | Grip strength in the dominant hand measured by a hydraulic dynamometer. Average of 3 measurements. | 1 | Physical function | 27.0 | 13.5 lbs‡ |

*SD = standard deviation; ROM = range of motion; ADL = activities of daily living; NA = not applicable; P-ROM = photographic range of motion; GVHD = graft-versus-host disease; FACTG = Functional Assessment of Cancer Therapy-General; QOL = quality of life; SF36 = Short Form 36; MCS = mental component score; PCS = physical component score; HAP = Human Activities Profile; MAS = maximum activity score; AAS = adjusted activity score.

†Derived from original design.

‡Derived from half of standard deviation of baseline values

## Table 2

Patient characteristics*

| | Joint/fascia manifestations at enrollment | | |
|---|---|---|---|
| | **Present (n = 164)** | **Absent (n = 403)** | ***P*** † |
| Time from transplantation to enrollment, median (IQR) months | 18 (11-25) | 11 (7-16) | < 0.001 |
| Case type, no. (%) | | | 0.001 |
| Incident | 79 (48) | 257 (64) | |
| Prevalent | 85 (52) | 146 (36) | |
| Patient age, median (IQR) years | 52 (42-58) | 51 (42-60) | 0.67 |
| Patient <18 years old, no. (%) | 4 (2) | 10 (2) | 0.98 |
| Patient gender, no. (%) | | | |
| Male | 94 (57) | 232 (58) | 0.96 |
| Female | 70 (43) | 171 (42) | |
| Patient race, no. (%) | | | 0.29 |
| White | 147 (90) | 363 (90) | |
| Non-white | 14 (8) | 38 (9) | |
| Unknown | 3 (2) | 2 (1) | |
| Stem cell source, no. (%) | | | 0.76 |
| Bone marrow | 10 (6) | 28 (7) | |
| Mobilized blood cells | 145 (89) | 358 (89) | |
| Cord blood | 9 (5) | 17 (4) | |
| Donor-patient gender combination, no. (%) | | | 0.69 |
| Female to male | 44 (27) | 120 (30) | |
| Other | 118 (72) | 280 (69) | |
| Not available | 2 (1) | 3 (1) | |
| HLA and donor type, no. (%) | | | 0.32 |
| HLA-matched related | 76 (46) | 164 (41) | |
| HLA-matched unrelated | 68 (42) | 168 (41) | |
| HLA-mismatched | 20 (12) | 69 (17) | |
| Not available | 0 (0) | 2 (1) | |
| Conditioning regimen, no. (%) | | | 0.009 |
| Myeloablative with high-dose TBI | 70 (43) | 116 (29) | |
| Non-myeloablative / reduced-intensity with low-dose TBI | 46 (28) | 135 (33) | |
| Without TBI | 46 (28) | 150 (37) | |
| Unknown | 2 (1) | 2 (1) | |
| Prior grades II-IV acute GVHD, no. (%) | | | 0.17 |
| Present | 92 (56) | 251 (63) | |
| Absent | 72 (44) | 152 (37) | |

* IQR = interquartile range; HLA = human leukocyte antigen; TBI = total body irradiation; GVHD = graft-versus-host disease.

† Two-sample t-test or Chi-square test of independence.

## Table 3

Chronic GVHD characteristics at enrollment*

| | Joint/fascia manifestations at enrollment | | | | |
|---|---|---|---|---|---|
| | N | Present (n = 164) | N | Absent (n = 403) | *P* † |
| NIH joint/fascia score, median (IQR) | 164 | 1 (1-2) | 403 | 0 (0-0) | NA |
| Hopkins fascia score, median (IQR) | 164 | 1 (0-1) | 403 | 0 (0-0) | < .001 |
| P-ROM total score, median (IQR) | 98 | 23 (21-24) | 231 | 25 (25-25) | < .001 |
| Other site involvement, no. (%) | | | | | |
| Skin | 164 | 123 (75) | 403 | 226 (56) | < 0.001 |
| Skin sclerosis | 164 | 86 (52) | 403 | 38 (9) | < 0.001 |
| Eye | 164 | 87 (53) | 403 | 189 (47) | 0.18 |
| Mouth | 164 | 78 (48) | 403 | 263 (65) | < 0.001 |
| Liver | 161 | 69 (43) | 402 | 221 (55) | 0.01 |
| Gastrointestinal tract | 164 | 54 (33) | 403 | 123 (31) | 0.58 |
| Lung | 164 | 79 (48) | 403 | 210 (52) | 0.40 |
| Genital tract | 147 | 21 (14) | 376 | 35 (9) | 0.10 |
| NIH global score, no. (%) | 164 | | 403 | | < 0.001 |
| Mild | | 4 (2) | | 49 (12) | |
| Moderate | | 82 (50) | | 211 (53) | |
| Severe | | 78 (48) | | 143 (35) | |
| Symptom measure, median (IQR) | | | | | |
| Lee muscle/joint subscale | 145 | 7 (4-11) | 336 | 3 (1-6) | < 0.001 |
| Lee overall symptom score | 145 | 24.3 (14.4-33.8) | 338 | 18.7 (11-28.6) | < 0.001 |
| 10-point overall global rating | 141 | 4 (3-6) | 328 | 3.5 (2-5) | < 0.001 |
| QOL measures, median (IQR) | | | | | |
| FACT-G | 139 | 76 (62-87) | 322 | 81 (69-90.3) | 0.003 |
| SF36-MCS | 137 | 47 (38-55.2) | 317 | 51 (40.8-55.9) | 0.06 |
| SF36-PCS | 137 | 37 (31.1-43.5) | 317 | 40 (32.2-47.9) | 0.002 |
| Physical function measures, median (IQR) | | | | | |
| HAP-MAS | 140 | 73 (61-82) | 326 | 73 (62-82) | 0.74 |
| HAP-AAS | 140 | 63 (51-74) | 326 | 62 (48-73) | 0.57 |
| Walk test (feet) | 139 | 482 (415-568) | 341 | 500 (404-575) | 0.84 |
| Grip test (lb) | 157 | 55.8 (40-81.7) | 377 | 61 (42.3-79.7) | 0.71 |

*IQR = interquartile range; NIH = National Institutes of Health; NA = not applicable; QOL = quality of life; FACT-G = Functional Assessment of Cancer Therapy-General; SF36 = hort Form 36; MCS = mental component score; PCS = physical component score; HAP = Human Activities Profile; MAS = maximum activity score; AAS = adjusted activity score.

†Two-sample t-test or Chi-square test of independence.