



Published in final edited form as:

Bone Marrow Transplant. 2014 April ; 49(4): 532–538. doi:10.1038/bmt.2013.225.

Investigator feedback about the 2005 NIH diagnostic and scoring criteria for chronic GVHD

Yoshihiro Inamoto¹, Madan Jagasia², William A. Wood³, Joseph Pidala⁴, Jeanne Palmer⁵, Nandita Khera⁶, Daniel Weisdorf⁷, Paul A. Carpenter¹, Mary E.D. Flowers¹, David Jacobsohn⁸, Paul J. Martin¹, Stephanie J. Lee¹, and Steven Z. Pavletic⁹ on behalf of the Chronic GVHD Consortium

¹Fred Hutchinson Cancer Research Center, Seattle, WA

²Vanderbilt University Medical Center, Nashville, TN

³Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁴H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL

⁵Medical College of Wisconsin, Milwaukee, WI

⁶Mayo Clinic Arizona, Scottsdale, AZ

⁷University of Minnesota Medical Center, Minneapolis, MN

⁸Children's National Medical Center, Center for Cancer and Blood Disorders, Washington, DC

⁹National Cancer Institute, Bethesda, MD

Abstract

The 2005 National Institutes of Health (NIH) consensus criteria for chronic graft-versus-host disease (cGVHD) have set standards for reporting. Many questions, however, have arisen regarding implementation and utilization. To identify perceived areas of controversy, we conducted an international survey on diagnosis and scoring of cGVHD. Agreement was observed for 50% to 83% of the 72 questions in 7 topic areas. There was agreement in the need for modifying criteria in 6 situations: 2 or more distinctive manifestations should be enough to diagnose cGVHD, symptoms not due to cGVHD should be scored differently, active disease and fixed deficits should be distinguished, a minimum threshold body surface area of hidebound skin involvement should be required for a skin score 3, asymptomatic oral lichenoid changes should be considered a score 1, and lung biopsy should be unnecessary to diagnose cGVHD in a patient with bronchiolitis obliterans as the only manifestation. The survey also identified 26 points of controversy. Whenever possible, studies should be conducted to confirm the appropriateness of

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence and reprint requests: Yoshihiro Inamoto, MD, PhD, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, D5-290, Seattle, WA 98109-1024, United States; telephone number: 206-667-2898; fax number: 206-667-1034; yinamoto@fhcrc.org.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Supplementary Information accompanies the paper on Bone Marrow Transplantation website (<http://www.nature.com/bmt>)

any revisions. In cases where data are not available, clarification of the NIH recommendations by consensus is necessary. This survey should inform future research in the field and revisions of the current consensus criteria.

Keywords

Allogeneic hematopoietic cell transplantation; chronic graft-versus-host disease; consensus; controversy; National Institutes of Health

INTRODUCTION

Chronic graft-versus-host disease (GVHD) is an immune-mediated disorder that occurs in 30–50% of patients after allogeneic hematopoietic cell transplantation (HCT).^{1–3} Chronic GVHD causes significant late morbidity and mortality and affects quality of life, survival and other transplant outcomes. In 2005, the National Institutes of Health (NIH) convened an expert conference to develop consensus on criteria for diagnosis, staging, pathology, biomarkers, response measurement, supportive care and design of clinical trials.^{4–9} The major goal of the consensus project was to develop a “standardized common language” among investigators focused on chronic GVHD in order to facilitate comparisons between clinical studies. While many studies have reported the validity of the Consensus criteria,^{10–27} implementation of the criteria in daily work has raised many practical questions.^{28, 29} Participants in the Conference anticipated that the advent of new data and experience applying the criteria to actual clinical situations may necessitate clarifications, corrections, refinement and perhaps revisions of the criteria.

The goals of this survey were to identify areas of confusion or disagreement with the NIH criteria particularly for diagnosis and severity scoring of chronic GVHD. We generated a questionnaire based on queries collected from various sources, and conducted a detailed survey among a voluntary group of international investigators who are interested in chronic GVHD research. The results of this survey should help to determine the next steps for research activities aiming to refine or modify the consensus criteria for chronic GVHD.

METHODS

Survey

Since the 2005 NIH Consensus Conference, the authors have collected frequently asked questions (FAQs) about the NIH consensus recommendations from various sources. Most FAQs came from health care professionals throughout the world, and some arose in the context of clinical trials. Two of the authors (Y.I. and S.J.L.) reviewed all collected FAQs and summarized them into 72 questions (Supplementary Appendix). Questions were formatted to elicit opinions about best practices for the future rather than asking how respondents are applying the current NIH recommendations. Two other authors (M.J. and S.P.) critically reviewed the draft question list and took the survey to establish the anticipated completion time before the questionnaire was finalized. An invitation to complete the survey was sent to 64 members of the International Chronic GVHD Special

Interest Group in February 2013, with two e-mail reminders sent 2 weeks and 3 weeks after the survey invitation. The International Chronic GVHD Special Interest Group is a voluntary group of investigators who are interested in chronic GVHD research. The group is organized by one of the authors (S.J.L.) and anyone can participate in the group by emailing their interest to chronicGVHDstudies@fhcrc.org. A web-based survey (Survey Monkey) was used to collect responses over a 5 week period. The invitation included information about a chance to win two \$500 gift cards for respondents who completed the survey by the deadline. The study protocol was approved by the Institutional Review Board of the Fred Hutchinson Cancer Research Center.

Definition of agreement in survey answers

Agreement for each question was considered high when 80% respondents chose the same answer, and was considered moderate when 60% but <80% respondents chose the same answer. Agreement was considered controversial when <60% respondents chose the same answer. Three questions that had multiple components (Q10, Q14, Q19) are reported according to the agreement level for each component.

Analyzed topic areas

Analysis was performed according to seven clinically relevant topic areas in the questionnaire (Table 1). Fourteen questions asked specifically whether the current NIH recommendations should be revised.

1. What are diagnostic (pathognomonic) and distinctive (suggestive of, but requiring histopathologic or additional testing for confirmation) criteria for chronic GVHD (21 questions)
2. Can pathology discriminate chronic GVHD from other causes? (6 questions)
3. Is biopsy necessary to diagnose chronic GVHD in certain organs? (5 questions)
4. How should severity of chronic GVHD be scored? (23 questions)
5. Should manifestations not due to GVHD be included in the scoring? (8 questions)
6. Some organs are scored in multiple ways. How should discrepancies between different evaluations be handled? (8 questions)
7. When the current NIH recommendations lack clarity or are silent about particularly clinical situations, how should they be scored? (8 questions)

RESULTS

Survey response and characteristics of participants

A total of 48 (75%) of 64 invited investigators completed the survey within 5 weeks after the survey was opened. Six (13%) respondents had 3–5 years of experience in caring for patients with chronic GVHD, 14 (29%) had 6–10 years, 15 (31%) had 11–20 years, and 11 (23%) had more than 20 years of experience. Forty-four (92%) of respondents were physicians and 40 (83%) considered themselves as experts in management of chronic GVHD. Respondents were from North America (67%), Europe (15%), South America

(10%), Asia (6%) and the Middle East (2%). The raw results of the survey are provided in Supplementary Appendix.

Agreement in answers according to topic areas

At least moderate agreement was observed for 50% to 83% of questions according to topic areas (Figure 1). High agreement was most frequently observed in responses to questions about the method of scoring symptoms not due to GVHD, where respondents clearly favored distinguishing symptoms based on whether they were attributable to chronic GVHD. Controversies were most frequently observed in responses to questions prompted by lack of clarity in the current NIH consensus, where respondents seemed to apply their own interpretations to the presented scenarios. Detailed results according to topic areas are shown in Table 1.

Opinions regarding revision of the current NIH recommendations

Among the 14 questions that asked whether the current NIH recommendations should be revised (Table 1), respondents agreed that recommendations should be revised in 6 points and should not be revised in 4 points. Opinions were controversial for 4 questions. The 6 points that respondents agreed with needing revision included the following:

High agreement

- 1 Two or more distinctive manifestations should be diagnostic of chronic GVHD.
- 2 Symptoms clearly not due to GVHD should be scored differently.
- 3 Active disease and fixed deficits should be distinguished when organ severity is scored.

Moderate agreement

- 4 A minimum threshold of body surface area involvement is required to score a skin severity of 3 for hidebound skin changes.
- 5 The mouth score should be score 1 if asymptomatic patients have diagnostic signs.
- 6 Clinical bronchiolitis obliterans syndrome without lung biopsy should be considered a diagnostic manifestation.

Summaries of answers for each topic

Topic 1: What are diagnostic and distinctive criteria for chronic GVHD?—

Respondents agreed that two or more distinctive manifestations should be sufficient for the diagnosis of chronic GVHD, that only one site with acute manifestations is enough to put a patient with chronic GVHD in the “overlap” category, and that “overlap” chronic GVHD and “progressive onset” are not interchangeable terms. Respondents agreed that the gastrointestinal (GI) and hepatic manifestations typically observed with acute GVHD are not distinctive for chronic GVHD. Respondents agreed that ocular dryness very early after transplantation should not be diagnosed as acute or chronic GVHD, and that clinical bronchiolitis obliterans syndrome (i.e., documented by pulmonary function tests and image

studies with negative work up for pathogens) should be considered a diagnostic manifestation without the need for confirmation by lung biopsy.

Many issues related to diagnostic and distinctive manifestations were controversial. For example, whether there were any diagnostic features for the eyes, liver and GI tract other than esophagus, whether classification of late acute GVHD and overlap chronic GVHD should be determined only by the current condition or by the history of chronic GVHD, whether gingivitis, oral mucositis and pain should remain as common signs seen in both acute and chronic GVHD, and whether joint pain mimicking rheumatoid arthritis is a manifestation of GVHD. Controversy also arose in the question of whether cryptogenic organizing pneumonia should remain a common sign of GVHD, should be a distinctive sign of chronic GVHD, or should be removed as a manifestation of GVHD.

Topic 2: Can pathology discriminate chronic GVHD from other causes?—

Respondents agreed that pathology was able to distinguish chronic GVHD from acute GVHD or other etiologies in the skin, lung, mouth and muscle but not in the liver and GI tract other than esophagus. Usefulness of pathology in distinguishing chronic GVHD in the fascia, esophagus and genital tract was controversial. We did not ask whether pathology could distinguish different infections.

Topic 3: Is biopsy necessary to diagnose chronic GVHD in certain organs?—

Respondents agreed that biopsy was not always necessary to confirm chronic GVHD in the skin, liver and GI tract if the patient had diagnostic chronic GVHD in other sites. Respondents agreed that diarrhea should be scored as chronic GVHD even for negative biopsy results in patients with chronic GVHD. It was controversial whether gastrointestinal biopsy was necessary to confirm GVHD in patients with upper GI symptoms.

Topic 4: Should criteria for severity scoring of chronic GVHD be revised?—

Respondents agreed that fixed deficits and active chronic GVHD should be distinguished when manifestations are scored, but it was controversial whether revisions should be made for both severity scoring and response measurement or only for response measurement. The current consensus does not clearly specify whether manifestations other than the eight core sites (skin, mouth, eyes, lung, liver, GI tract, joints and fascia and genital tract) should be included in the global scoring system. Respondents agreed esophageal stricture, thrombocytopenia, and pericardial or pleural effusion but not cardiac complications should be included, although they remain difficult to score. Inclusion of performance status and other manifestations was controversial. For the skin, respondents agreed that maculopapular rash, lichenoid features, erythroderma, sclerosis, erythema, papulosquamous lesions or ichthyosis, poikiloderma, keratosis pilaris and hair involvement should all be considered when determining the skin score, while controversy arose in the question of whether hypo or hyper-pigmentation, pruritus and nail involvement should be considered when determining the skin score as currently recommended. Respondents agreed that hidebound changes should not be rated as a skin score 3 if the body surface area affected was minimal, and that the mouth score 1 should be revised so that it includes asymptomatic lichenoid changes that are currently scored as 0. Respondents reaffirmed the current NIH recommendations that mechanical interventions for dry eye such as punctual plugging and corneal lenses, and

visual impairment should be considered in the eye severity score. The use of a dilator for vaginal stricture in the absence of symptoms as an indicator of severe involvement was controversial.

Topic 5: Should manifestations not due to GVHD be included in the scoring?

—Overall, respondents agreed that symptoms clearly attributable to causes other than GVHD should not be scored. For example, respondents agreed that eye, joint or fascia and lung manifestations should not be scored if they predated transplantation and were stable. In cases where clinicians are not sure of the etiology, respondents agree that symptoms should be included in the chronic GVHD scoring. It was controversial whether the score should be downgraded by a 1 point from the actual score when other proven etiologies were present.

Topic 6: How should discrepancies between different evaluations be handled?

—For the skin, eyes, lung and genital tract, respondents agreed with the use of the worst manifestation when the patient had discrepant findings in different areas rather than using the lesser or average score. For the lung, respondents agreed that pulmonary symptoms should be scored 0 if pulmonary function tests were normal, although the current consensus recommended taking the higher of the symptom and pulmonary function test score. Answers for the mouth and liver were controversial.

Topic 7: When the current NIH recommendations lack clarity or are silent about particularly clinical situations, how should they be scored?

—Despite lack of clear guidance in some of the consensus recommendations, agreement in opinions was observed for half of the questions. Respondents agreed that erectile dysfunction is not a GVHD manifestation, that some eye drops such as antibiotics and glaucoma medications should not be included for the eye score based on the frequency of eye drop use, and that all liver function tests including serum bilirubin, alanine aminotransferase, aspartate aminotransferase and alkaline phosphatase should be considered in staging late acute GVHD of the liver. Controversies were related to duration of resolution required for the “recurrent” GVHD category, the reference time point used for calculation of weight loss as a manifestation of GI involvement, and the methods used to evaluate patients when pulmonary function tests or gynecological exam results were missing.

DISCUSSION

In the past, studies of chronic GVHD have been compromised by lack of standardized diagnosis, grading and response assessment of chronic GVHD. The 2005 NIH Consensus Conference provided a common language and set of criteria, but clinical experience with the recommendations has exposed some areas of ambiguity that need clarification. Although agreement rates varied according to topic areas, at least moderate agreement was observed for 50% of questions addressing these areas, suggesting that consensus can be reached on these issues.

For the most part, the controversial areas would not substantially change interpretation of the scoring or response criteria. However, two main areas where respondents disagreed with the current NIH recommendations (the rules for scoring symptoms not due to GVHD and

Author Manuscript

Author Manuscript

Author Manuscript

lack of distinction between active disease and fixed deficits) might have a major impact on scoring depending on the number of other conditions contributing to the organ abnormalities. This question was discussed during the 2005 Consensus Conference, and it was decided that incorporating considerations of attribution or reversibility would complicate severity scoring as no reliable definitions of reversibility existed. The results of our survey show that real-world experience has led to discontent with the inability to include these considerations when scoring severity. Future studies should focus on clarifying these areas as they are critical to both severity scoring and response measurement. To account for manifestations not due to GVHD, scores may be down-graded in a manner similar to the modified scoring for acute GVHD.³⁰ Since such an adjustment for concomitant diseases might result in increased heterogeneity in severity grading, this approach requires validation. To account for the distinction between activity and fixed deficits, an activity score for chronic GVHD could be developed, similar to the Systemic Lupus Erythematosus Disease Activity Index³¹ or Crohn's Disease Activity Index.³² Identification and validation of biomarkers to distinguish active and inactive chronic GVHD would be very useful since these situations may be indistinguishable by physical exam.

Other issues where respondents disagreed with current NIH recommendations are the need for lung biopsy to diagnose bronchiolitis obliterans when they do not have other manifestations of chronic GVHD. Many clinicians recognize the risk and difficulty of lung biopsy for diagnosis of bronchiolitis obliterans and are not willing to subject patients to such risks. Similar topics were discussed at the 2009 European meeting²⁹ and at the 2012 BMT Tandem Meetings.²⁸

Author Manuscript

Another disagreement is with the rule for scoring the mouth as a 0 for asymptomatic lichenoid changes. This was based on the assumption that activities of daily living would not be affected by asymptomatic oral lichenoid changes and therefore they should be scored as a 0 in severity assessment. In addition, systemic treatment is often not indicated if oral lichenoid changes are the sole manifestation of chronic GVHD. On the other hand, if isolated lichenoid changes are not captured in organ scoring, the association of these changes with long-term sequela including secondary oral cancer would be missed.³³ Directed studies are warranted to determine the appropriateness of modifying the rule for scoring the mouth, since changes in the current mouth score correlated well with both clinician and patient-perceived response in oral GVHD.²⁶

Author Manuscript

Author Manuscript

Many of the controversial issues were identified because they are common situations in clinical practice but the current NIH recommendations are unclear or lack guidance. For example, "recurrent" late acute GVHD is defined as development of acute GVHD beyond day 100 after prior acute GVHD has resolved, but the number of GVHD-free days required to be considered "recurrent" versus "persistent" has never been defined. Another example is the chronic GVHD GI severity score, which is based in part on the percentage of weight loss, although the reference time point is not specified. A routine gynecological examination for patients with chronic GVHD is required, since manifestations of genital GVHD are not always reported and are not captured unless the physical exam includes the genitalia. On the other hand, rules for scoring in the absence of gynecological examination should be given. The results of the survey also highlight the need for developing a consensus about

histopathology and biomarkers of liver GVHD since clinical and laboratory findings are not diagnostic. Given the current lack of empirical data, evidence-based resolution is not possible. Efforts should be made to design appropriate studies to address those questions. Provisional consensus could be reached in other cases where need for clarifications is urgent or obvious.

This study has some limitations. First, results were derived from a relatively small group of committed investigators. Mitigating this concern is that respondents are from various international regions, half have more than 10 years' experience of caring patients with chronic GVHD, many report they use the NIH criteria in clinical practice, and most have extensive experience with clinical trials in chronic GVHD. Thus, the results of this study should provide meaningful opinions. Second, we could not evaluate differences in answers according to investigators' regions or experience including years in practice and the number of patients they have seen because of limited numbers of participants.

In summary, this survey highlights areas of controversies that were not anticipated by the NIH criteria or where additional clinical experience has led investigators to question the original recommendations. These results will be useful guidance for revisiting the NIH consensus criteria, and a conference is planned in June 2014. In situations where sufficient data are available, revisions of the current recommendations will be made. In situations where additional information is needed, studies should be performed to collect these data. For other situations where data cannot be generated and recommendations are truly based on opinions, the goal should be to achieve a clear consensus to ensure standardized use of the criteria.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank all participants in the web survey for their valuable time and opinions. This work was supported by grants CA118953 and CA163438 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. Lee SJ, Vogelsang G, Flowers ME. Chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2003; 9:215–33. [PubMed: 12720215]
2. Wolff D, Gerbitz A, Ayuk F, Kiani A, Hildebrandt GC, Vogelsang GB, et al. Consensus conference on clinical practice in chronic graft-versus-host disease (GVHD): first-line and topical treatment of chronic GVHD. *Biol Blood Marrow Transplant.* 2010; 16:1611–28. [PubMed: 20601036]
3. Wolff D, Schleuning M, von Harsdorf S, Bacher U, Gerbitz A, Stadler M, et al. Consensus Conference on Clinical Practice in Chronic GVHD: Second-Line Treatment of Chronic Graft-versus-Host Disease. *Biol Blood Marrow Transplant.* 2011; 17:1–17. [PubMed: 20685255]
4. 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]
5. Shulman HM, Kleiner D, Lee SJ, Morton T, Pavletic SZ, Farmer E, et al. Histopathologic diagnosis of chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: II. Pathology Working Group Report. *Biol Blood Marrow Transplant*. 2006; 12:31–47. [PubMed: 16399567]
 6. Schultz KR, Miklos DB, Fowler D, Cooke K, Shizuru J, Zorn E, et al. Toward biomarkers for chronic graft-versus-host disease: National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: III. Biomarker Working Group Report. *Biol Blood Marrow Transplant*. 2006; 12:126–37. [PubMed: 16443511]
 7. 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]
 8. Couriel D, Carpenter PA, Cutler C, Bolanos-Meade J, Treister NS, Gea-Banacloche J, et al. Ancillary therapy and supportive care of chronic graft-versus-host disease: national institutes of health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: V. Ancillary Therapy and Supportive Care Working Group Report. *Biol Blood Marrow Transplant*. 2006; 12:375–96. [PubMed: 16545722]
 9. Martin PJ, Weisdorf D, Przepiora D, Hirschfeld S, Farrell A, Rizzo JD, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: VI. Design of Clinical Trials Working Group report. *Biol Blood Marrow Transplant*. 2006; 12:491–505. [PubMed: 16635784]
 10. Jagasia M, Giglia J, Chinratanalab W, Dixon S, Chen H, Frangoul H, et al. Incidence and outcome of chronic graft-versus-host disease using National Institutes of Health consensus criteria. *Biol Blood Marrow Transplant*. 2007; 13:1207–15. [PubMed: 17889358]
 11. Arora M, Nagaraj S, Witte J, DeFor TE, MacMillan M, Burns LJ, et al. New classification of chronic GVHD: added clarity from the consensus diagnoses. *Bone Marrow Transplant*. 2009; 43:149–53. [PubMed: 18794869]
 12. Cho BS, Min CK, Eom KS, Kim YJ, Kim HJ, Lee S, et al. Feasibility of NIH consensus criteria for chronic graft-versus-host disease. *Leukemia*. 2009; 23:78–84. [PubMed: 18830253]
 13. 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]
 14. Thepot S, Zhou J, Perrot A, Robin M, Xhaard A, de Latour RP, et al. The graft-versus-leukemia effect is mainly restricted to NIH-defined chronic graft-versus-host disease after reduced intensity conditioning before allogeneic stem cell transplantation. *Leukemia*. 2010; 24:1852–8. [PubMed: 20827288]
 15. Sato T, Ichinohe T, Kanda J, Yamashita K, Kondo T, Ishikawa T, et al. Clinical significance of subcategory and severity of chronic graft-versus-host disease evaluated by National Institutes of Health consensus criteria. *Int J Hematol*. 2011; 93:532–41. [PubMed: 21465116]
 16. Arai S, Jagasia M, Storer B, Chai X, Pidala J, Cutler C, et al. Global and organ-specific chronic graft-versus-host disease severity according to the 2005 NIH Consensus Criteria. *Blood*. 2011; 118:4242–9. [PubMed: 21791424]
 17. Pidala J, Kurland B, Chai X, Majhail N, Weisdorf DJ, Pavletic S, et al. Patient reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. *Blood*. 2011; 117:4651–4657. [PubMed: 21355084]
 18. Pidala J, Kim J, Anasetti C, Nishihori T, Betts B, Field T, et al. The global severity of chronic graft-versus-host disease, determined by National Institutes of Health consensus criteria, is associated with overall survival and non-relapse mortality. *Haematologica*. 2011; 96:1678–84. [PubMed: 21791465]
 19. Mitchell SA, Jacobsohn D, Thormann Powers KE, Carpenter PA, Flowers ME, Cowen EW, 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; 17:619–29.
20. 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]
 21. Del Fante C, Scudeller L, Viarengo G, Bernasconi P, Perotti C. Response and survival of patients with chronic graft-versus-host disease treated by extracorporeal photochemotherapy: a retrospective study according to classical and National Institutes of Health classifications. *Transfusion*. 2012; 52:2007–15. [PubMed: 22320219]
 22. 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]
 23. Terwey TH, Le Duc TM, Hemmati PG, le Coutre P, Nagy M, Martus P, et al. NIH-defined graft-versus-host disease and evidence for a potent graft-versus-leukemia effect in patients with acute lymphoblastic leukemia. *Ann Oncol*. 2012 (in press).
 24. Inamoto Y, Martin PJ, Chai X, Jagasia M, Palmer J, Pidala J, et al. Clinical Benefit of Response in Chronic Graft-versus-Host Disease. *Biol Blood Marrow Transplant*. 2012; 18:1517–24. [PubMed: 22683612]
 25. Palmer J, Lee SJ, Chai X, Storer B, Flowers ME, Schultz K, et al. Poor Agreement between Clinician Response Ratings and Calculated Response Measures in Patients with Chronic Graft-versus-host Disease. *Biol Blood Marrow Transplant*. 2012; 18:1649–1655. [PubMed: 22691695]
 26. Treister N, Chai X, Kurland B, Pavletic S, Weisdorf D, Pidala J, et al. Measurement of oral chronic GVHD: results from the Chronic GVHD Consortium. *Bone marrow transplantation*. 2013
 27. Aisa Y, Mori T, Kato J, Yamane A, Kohashi S, Kikuchi T, et al. Validation of NIH consensus criteria for diagnosis and severity-grading of chronic graft-versus-host disease. *Int J Hematol*. 2013; 97:263–71. [PubMed: 23371545]
 28. Blazar B, White ES, Couriel D. Understanding chronic GVHD from different angles. *Biol Blood Marrow Transplant*. 2012; 18:S184–8. [PubMed: 22226105]
 29. Greinix HT, Loddenkemper C, Pavletic SZ, Holler E, Socie G, Lawitschka A, et al. Diagnosis and staging of chronic graft-versus-host disease in the clinical practice. *Biol Blood Marrow Transplant*. 2011; 17:167–75. [PubMed: 20691801]
 30. Przepiorka D, Weisdorf D, Martin P, Klingemann HG, Beatty P, Hovs J, et al. 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant*. 1995; 15:825–8. [PubMed: 7581076]
 31. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. *Arthritis Rheum*. 1992; 35:630–40. [PubMed: 1599520]
 32. Best WR, Becktel JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology*. 1976; 70:439–44. [PubMed: 1248701]
 33. Curtis RE, Rowlings PA, Deeg HJ, Shriner DA, Socie G, Travis LB, et al. Solid cancers after bone marrow transplantation. *N Engl J Med*. 1997; 336:897–904. [PubMed: 9070469]

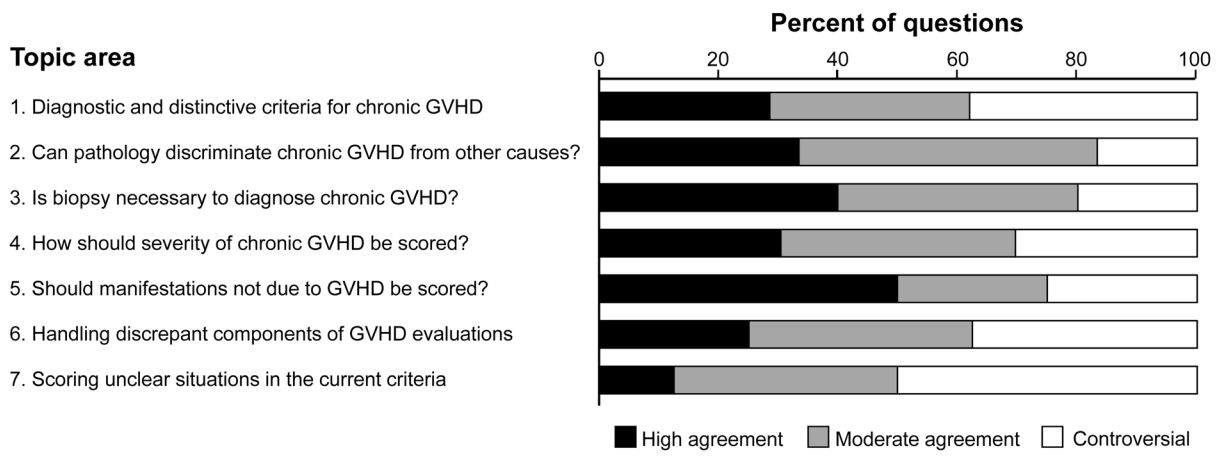


Figure 1.
Agreement in answers according to topic areas.

Table 1

Agreement level according to topic areas

Topic 1: What are diagnostic and distinctive criteria for chronic GVHD?		
Agreement level	Agreed answer or controversial question	Category
High* (R)	Two or more distinctive manifestations should be considered sufficient to diagnose chronic GVHD. (Q3)	Diagnosis
High	Only one site with acute manifestations (skin, liver or GI) is enough to diagnose "overlap" chronic GVHD. (Q5)	Subcategory of GVHD
High	Overlap chronic GVHD and progressive onset are NOT interchangeable terms. (Q8)	Subcategory of GVHD
High	It is difficult to distinguish deep from superficial sclerosis in the abdominal skin among obese patients. (Q23)	Skin
High	By careful history taking and physical examination of fascia and joints, you can distinguish joint problems related to chronic GVHD from other cause of joint impairment. (Q53)	Joint/fascia
High	Nephrotic syndrome after allogeneic transplantation should be considered a manifestation of chronic GVHD. (Q72)	Other sites
Moderate* (NR)	There should be no distinctive features of chronic GVHD for the GI tract and liver. (Q2)	Diagnosis
Moderate* (NR)	There should be no pediatric modifications to the categorizations of diagnostic and distinct manifestations. (Q4)	Diagnosis
Moderate* (NR)	It is not necessary to revise the terms "acute" and "chronic" GVHD to something else. (Q6)	Subcategory of GVHD
Moderate	There are no imaging methods that can help to distinguish deep from superficial sclerosis. (Q24)	Skin
Moderate	Neither acute nor chronic GVHD should be diagnosed for a patient who had ocular dryness very early after transplantation (for example, 30 days after transplantation), with no other manifestations of chronic GVHD. (Q26)	Eye
Moderate	Isolated early fasciitis manifested by edema is diagnostic for chronic GVHD. (Q52)	Joint/fascia
Moderate* (R)	Clinical bronchiolitis obliterans syndrome should be considered a diagnostic manifestation (sufficient to make the diagnosis of chronic GVHD). (Q56)	Lung
Controversial*	Currently, there are no diagnostic features of chronic GVHD for the eyes, liver and GI tract except esophagus. Should there be? (Q1)	Diagnosis
Controversial	Do you categorize the following patient as late acute GVHD or overlap chronic GVHD? The patient had overlap chronic GVHD (oral lichenoid changes and gut GVHD). All manifestations were completely resolved after six months of systemic treatment. Now 2 years after transplantation, the patient has recurrent lower gut GVHD without any other signs of chronic GVHD. (Q7)	Subcategory of GVHD
Controversial	If a patient has "new ocular sicca documented by Schirmer test" or "a new onset of keratoconjunctivitis with Schirmer score <10 mm", is this sufficient to diagnose chronic GVHD? (Q33)	Eye
Controversial*	Should gingivitis, oral mucositis and pain continue to be considered "common" signs, even if mouth is not a recognized target organ in acute GVHD? Or should these be considered distinctive signs of chronic GVHD? (Q40)	Mouth
Controversial	Do you think joint pain mimicking rheumatoid arthritis after transplant is joint GVHD? (Q54)	Joint
Controversial	Is cryptogenic organizing pneumonia a form of lung GVHD? (Q57)	Lung
Controversial*	Should cryptogenic organizing pneumonia still be considered a "common" sign, since lung is not a recognized target organ in acute GVHD? (Q58)	Lung
Controversial	How do you determine if peripheral neuropathy is due to chronic GVHD in a patient with an established diagnosis of chronic GVHD? (Q71)	Other sites
Topic 2: Can pathology discriminate chronic GVHD from other causes?		
Agreement level	Agreed answer or controversial question	Category
High	Pathologists can NOT distinguish between acute and chronic GVHD in the liver. (Q10)	Pathology
High	Pathologists can NOT confidently diagnose liver chronic GVHD. (Q11)	Pathology

Topic 1: What are diagnostic and distinctive criteria for chronic GVHD?		
Agreement level	Agreed answer or controversial question	Category
Moderate	Pathologists can NOT distinguish in GI tract except for esophagus. (Q10)	Pathology
Moderate	Pathologists can distinguish between acute and chronic GVHD in the skin, lung and mouth. (Q10)	Pathology
Moderate	If muscle biopsy is positive for myositis but diagnostic or distinctive features of chronic GVHD are absent in other sites, it should be sufficient to diagnose chronic GVHD. (Q55)	Muscle
Controversial	Pathologists can distinguish between acute and chronic GVHD in fascia, esophagus, and genital tract. (Q10)	Pathology
Topic 3: Is biopsy necessary to diagnose chronic GVHD in certain organs?		
Agreement level	Agreed answer or controversial question	Category
High* (NR)	Skin biopsy is NOT mandatory for diagnosis of skin chronic GVHD. (Q16)	Skin
High	If a patient with already diagnosed chronic GVHD has LFT abnormalities but no liver biopsy, LFT abnormalities should be considered GVHD. (Q49)	Liver
Moderate	We should score diarrhea in the NIH GI scoring section when the patient has chronic GVHD in other sites but biopsy is negative for GI GVHD. (Q42)	GI
Moderate	Diarrhea should be scored as GI GVHD if no biopsy is done but a patient has diagnostic chronic GVHD in other sites. (Q43)	GI
Controversial	A patient has chronic GVHD in other sites plus nausea and anorexia. Is a biopsy required for the diagnosis of GI involvement? (Q44)	GI
Topic 4: How should severity of chronic GVHD be scored?		
Agreement level	Agreed answer or controversial question	Category
High	Cardiomyopathy, cardiac conduction defects, and coronary artery involvement should NOT be included in the global scoring system. (Q14)	Global score
High* (R)	We should revise the current consensus that recommends rating organ severity without distinguishing between active disease and fixed deficits. (Q15)	Response
High	Maclopopular rash, lichen planus-like feature, erythroderma, sclerotic features, erythema, papulosquamous lesions or ichthyosis and poikiloderma should be considered for calculating body surface area (BSA). (Q19)	Skin
High	Pruritus should NOT be considered for calculating BSA. (Q19)	Skin
High	Just pruritus (without any skin changes) is NOT sufficient for NIH skin score 1 or greater. (Q21)	Skin
High	A patient had punctal plugging and had symptomatic relief to such an extent that he requires eye drops only 2 times a day. If you confirm that the punctal plugs fall out, the NIH eye score should be score 1. (Q29)	Eye
High	If a patient lost vision in one eye because of chronic GVHD but is completely asymptomatic in the other eye, the NIH eye score is score 3. (Q36)	Eye
Moderate	Esophageal stricture or web, thrombocytopenia, pericardial effusion, and pleural effusion should be included in the global scoring system. (Q14)	Global score
Moderate* (R)	The rule should be revised in scoring 3 for hidebound changes in only a small area (for example 1% of legs). (Q17)	Skin
Moderate	Keratosis pilaris and hair involvement should be considered for calculating BSA. (Q19)	Skin
Moderate	Just nail and/or hair involvement is sufficient for NIH skin score 1. (Q20)	Skin
Moderate	Just hyperpigmentation and/or hypopigmentation of skin is sufficient for NIH skin score 1 or greater. (Q22)	Skin
Moderate	A patient had punctal plugging and had symptomatic relief to such an extent that he requires eye drops only 2 times a day. If you confirm that the punctal plugs are still in the eyes, the NIH eye score should be score 2. (Q28)	Eye

Topic 1: What are diagnostic and distinctive criteria for chronic GVHD?		
Agreement level	Agreed answer or controversial question	Category
Moderate	A patient started special contact lenses for treatment of ocular GVHD and had symptomatic relief to such an extent that he requires eye drops 2 times a day. The NIH eye score should be score 3. (Q30)	Eye
Moderate* (R)	If a patient has diagnostic signs such as lichenoid changes but has no oral symptoms, the NIH mouth score should be score 1. (Q37)	Mouth
Moderate	We should consider superficial mucoceles that come and go when you determine the NIH mouth score. (Q41)	Mouth
Controversial*	Performance status scoring is not incorporated into the NIH global scoring system. Should this be revised? (Q13)	Global score
Controversial	Ascites, eosinophilia, polymyositis, nephrotic syndrome, myasthenia gravis should be included in the global scoring system. (Q14)	Global score
Controversial	Nail involvement, hypopigmentation and hyperpigmentation should be considered for calculating BSA. (Q19)	Skin
Controversial	Should we consider excessive tearing as one form of GVHD in the NIH eye score? (Q34)	Eye
Controversial	A female patient is asymptomatic due to sexual inactivity and has moderate signs of genital GVHD. What is the NIH genital score? (Q67)	Genital
Controversial	A female patient tells you that she is asymptomatic but uses a dilator for her fixed moderate vaginal stricture. What is the NIH genital score? (Q68)	Genital
Controversial	Is vaginal dryness sufficient for NIH genital score 1 or greater? (Q69)	Genital
Topic 5: Should manifestations not due to GVHD be included in the scoring?		
Agreement level	Agreed answer or controversial question	Category
High* (R)	We should revise the current consensus that recommends rating all symptoms even if you do not think the symptoms are due to GVHD. (Q12)	Overall
High	A patient has just been diagnosed with chronic GVHD in the mouth. If the patient has been using eye drops 3 times a day due to dry eye starting before transplant, the patient should NOT be scored for ocular chronic GVHD. (Q27)	Eye
High	Joint tightness due to prior injury or avascular necrosis should be scored as 0 in the NIH joint score. (Q51)	Joint/fascia
High	If a patient has chronic obstructive pulmonary disease (COPD) before transplant, we should determine the NIH PFT score according to post-transplant PFTs only if they worsen from pre-transplant PFTs. (Q62)	Lung
Moderate	A patient has mild loose stool and the colon biopsy is positive for GVHD. The patient also had 10% weight loss as compared to one month ago, but you attribute the weight loss to poorly controlled steroid-induced diabetes. The NIH GI score should be score 2. (Q46)	GI
Moderate	Dyspnea that you believe is due to steroid myopathy should be scored as 0. (Q59)	Lung
Controversy	A patient uses eye drops 2 times a day due to dry eye prior to transplant. After transplant, he is diagnosed with chronic GVHD and increases the frequency of eye drops to 4 times a day due to worsening eye dryness. How do you rate the NIH eye score? (Q31)	Eye
Controversial	If other etiologies are confirmed for liver abnormalities (for example, hemochromatosis, viral hepatitis, leukemia invasion, drug side effect, or alcohol consumption), how do you rate the NIH liver score for patients with chronic GVHD? (Q50)	Liver
Topic 6: How should discrepancies between different evaluations be handled?		
Agreement level	Agreed answer or controversial question	Category
High	When you see hyperpigmentation and lichenoid in the same area, the Total Skin Score (Vienna score) is score 2. (Q25)	Skin
High	If eye symptoms differ between left and right eyes, I use the worse eye for rating the NIH eye score. (Q35)	Eye
Moderate	When lung symptom scores differ from PFT scores, higher values are used for final lung scores. We should continue to use PFT scores to determine the lung score whenever PFT results are available even if the patient has no pulmonary symptoms. (Q60)	Lung

Topic 1: What are diagnostic and distinctive criteria for chronic GVHD?		
Agreement level	Agreed answer or controversial question	Category
Moderate	If a patient has pulmonary symptoms with normal PFT, we should score 0 for the NIH lung score, since you are not sure whether the symptoms are due to GVHD. (Q61)	Lung
Moderate	The NIH genital score is score 3 for a female who has mild dyspareunia and severe signs on gynecological exam. (Q70)	Genital tract
Controversial	If the patient has extensive oral lichenoid changes but has only mild symptoms, how should we rate the NIH mouth score? (Q39)	Mouth
Controversial	How should we rate the NIH mouth score for moderate oral sensitivities without lichenoid changes or other signs of chronic GVHD? (Q38)	Mouth
Controversial	A patient has normal AST, ALT and bilirubin, but has elevated alkaline phosphatase (AP). You do not have isozyme information for AP. How do you rate the NIH liver score? (Q48)	Liver
Topic 7: When the current NIH recommendations lack clarity or are silent about particularly clinical situations, how should they be scored?		
Agreement level	Agreed answer or controversial question	Category
High	Erectile dysfunction is NOT a symptom of genital GVHD. (Q64)	Genital
Moderate	All types of eye drops should NOT be included when counting the frequency of eye drops for rating the NIH eye score (i.e., also include cyclosporine eye drop, steroid eye drop, antibiotic eye drops)? (Q32)	Eye
Moderate	The same laboratory tests for NIH chronic GVHD (ie, ALT, AST, bilirubin and alkaline phosphatase) should be applied to staging of late acute GVHD in the liver. (Q47)	Liver
Moderate	We should score the genitals for men. (Q65)	Genital tract
Controversial	When you diagnose recurrent late acute GVHD or quiescent chronic GVHD, how many days of acute GVHD resolution are required before symptoms start again? (Q9)	Subcategory of GVHD
Controversial	What is the reference time point used for calculation of weight loss? (Q45)	GI
Controversial	If a patient doesn't have PFTs results, how should we determine the NIH lung score? (Q63)	Lung
Controversial	Can the NIH genital score be completed without gynecological exam? (Q66)	Genital

* Questions that asked specifically whether the current NIH recommendations should be revised. (R) = Revision recommended by respondents. (NR) = Respondents recommended against revision.