Published in final edited form as:

Best Pract Res Clin Haematol. 2014 ; 27(0): 283–287. doi:10.1016/j.beha.2014.10.011.

Improved accuracy of acute graft-versus-host disease staging among multiple centers

John E. Levine, MD, MS^a [Professor of Pediatrics], William J. Hogan, MD^b [Assistant Professor of Medicine], Andrew C. Harris, MD^a [Lecturer in Pediatrics], Mark R. Litzow, MD^b [Professor of Medicine], Yvonne A. Efebera, MD^c [Assistant Professor of Internal Medicine], Steven M. Devine, MD^c [Professor of Internal Medicine], Ran Reshef, MD^d [Assistant Professor of Medicine], and James L.M. Ferrara, MD, DSc^{a,*} [Ruth Heyn Professor of Pediatrics]

^aBlood and Marrow Transplantation Program, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109

^bBlood and Marrow Transplant Program, Mayo Clinic, 200 First St. SW, Rochester, MN 55905

^cBlood and Marrow Transplant Program, Ohio State University, 300 W. 10th Ave. Columbus, OH 43210

^dBlood and Marrow Transplant Program, University of Pennsylvania, 3400 Civic Center Blvd., Philadelphia PA 19104

Abstract

The clinical staging of acute graft-versus-host disease (GVHD) varies significantly among bone marrow transplant (BMT) centers, but adherence to long-standing practices poses formidable barriers to standardization among centers. We have analyzed the sources of variability and developed a web-based remote data entry system that can be used by multiple centers simultaneously and that standardizes data collection in key areas. This user-friendly, intuitive interface resembles an online shopping site and eliminates error-prone entry of free text with drop-down menus and pop-up detailed guidance available at the point of data entry. Standardized documentation of symptoms and therapeutic response reduces errors in grade assignment and allows creation of confidence levels regarding the diagnosis. Early review and adjudication of borderline cases improves consistency of grading and further enhances consistency among centers. If this system achieves widespread use it may enhance the quality of data in multicenter trials to prevent and treat acute GVHD.

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding author: James L.M. Ferrara, Tel.: +1 734 615 1339, Fax: +1 734 647 9271, ferrara@umich.edu. 10th Ave. Columbus, OH 43210.

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

Conflict of Interest Statement

Please provide; even if "None".

Keywords

Bone marrow transplant; BMT; data entry system; guidance; graft-versus-host disease; GVHD; multiple centers; online; staging; standardization; web

Introduction

The clinical staging of acute graft-versus-host disease (GVHD) is inconsistent among transplant centers and highly prone to errors [1]. The very definition of GVHD remains variable, with the cumulative incidence of GVHD grade II-IV ranging from 40% to 80% following T-cell replete bone marrow transplant (BMT) [2-4]. Although this poor concordance between transplant centers has been recognized for over 25 years [5,6], different practices of long standing among centers pose formidable barriers to harmonization. These differences become glaringly apparent during the conduct of multi-center trials when it can be very difficult to decide whether a patient actually experienced GVHD. This article reviews some new approaches to multicenter efforts to improve GVHD grading that have been piloted in a consortium led by investigators at the University of Michigan to address these common problems using remote data entry system and near-real time adjudication.

Standardized, user-friendly data collection reduces needless variations

Data collection systems at BMT centers possess many idiosyncratic features as a result of practices that evolved over years, but they are familiar to their users. Uniform data collection systems that are fast and easy-to-learn are highly desirable; investigators at the University of Michigan have recently developed flexible, intuitive, web-based interfaces similar to online shopping sites in order to standardize data collection among BMT centers. Entries are completed by radio button mouse clicks and drop down menus rather than free text entry. The first layer of the data collection asks yes/no questions, eg, “was there a rash this week?” If no, no further entry is needed. If yes, additional questions appear to determine etiology, extent, and treatment. As an example of a desirable feature, detailed guidance is available at the point of data entry, appearing only when clicked, which keeps the form visually uncluttered, but makes reference to guidance easy. Logic checks warn users of missing or inconsistent data via pop-up alerts much the same way online shoppers are warned of incorrect credit card entries.

In our experience, physicians who are familiar with the data entry system are better equipped to troubleshoot problems, and thus both physicians and data managers were encouraged to participate in a data entry webinar where a lead data coordinator enters data from source documents while answering questions posed by participants. Following the webinar, data managers from each center entered data in an electronic “sandbox” populated with test patients. The time requirements to collect the data appear to be manageable. When patients have no new symptoms, data entry can be completed in less than 1 minute; when new symptoms appear, the forms can still be completed in fewer than 5 minutes. Feedback from centers involved in the testing has been very positive.

Documentation of granular details improves consistency of GVHD staging

A 55% body surface rash is stage 3 skin GVHD, but “rash” does not distinguish active inflammatory erythema from inactive hyperpigmentation; thus, the above rash may be categorized by different observers either as a stage 2 or a stage 3, changing the overall grade and the need to treat with systemic steroids. Gastrointestinal (GI) staging is even more prone to inaccuracies because it requires accurate measurement of diarrhea volume. When diarrhea starts at home, patients almost always quantify by episodes, not volume, and estimation of volume based on history alone is inherently flawed. “Five episodes of diarrhea” can be staged as stage 1 or 2, *again changing the overall grade*. In order to address this type of problem, we created standardized GVHD guidance after a review of weekly GVHD grades that identified both common and uncommon sources of confusion. This guidance addresses important but vexing scenarios such as staging GI GVHD that develops at home. We reviewed hospital flow-sheets for 300 patients at the University of Michigan with GI GVHD containing both the total daily volume of diarrhea and the number of episodes. From these sources we calculated the average volume per diarrhea episode as 200 mL/episode or 3 mL/kg for children <50 kg. We use this guidance to estimate diarrhea when only the number of episodes are available (approximately 30% of GI GVHD at onset). Three different centers tested the guidance on a total of 75 GVHD cases for consistency, clarity, and utility and found it both efficient and easy to follow.

Direct assessments reduce errors in grade assignment

GVHD staging is often performed by residents, fellows, or physician extenders with limited GVHD experience. Thus, a *correctly assessed* 40% skin rash can be *incorrectly interpreted* as stage 3 (overall grade II) instead of stage 2 (overall grade I). Furthermore, raw data such as diarrhea is usually updated daily in electronic progress notes, but GVHD assessments are copied and pasted, creating potential internal inconsistencies *within a single progress note*. Recognition and correction of such errors requires careful evaluation of the raw data by an experienced clinician, but this time-sensitive, labor-intensive step is impractical. Once incorrect GVHD stages are transcribed, *post facto* correction is extremely difficult and rarely successful. We have attempted to avoid this problem by collecting raw data (extent of rash, volume of diarrhea) rather than the GVHD stages. Drop down menus can permit selection of units used in the medical record (eg, steroid doses in mg/day or mg/kg), simplifying data entry and avoiding mathematical errors. Any clinical staging system (eg, modified Glucksberg [7], IBMTR [8]) can then be easily applied to the raw data. These data are collected weekly through day 100 or resolution of symptoms, whichever comes later. We collect the same information for GVHD that develops after day 100 during the first 4 weeks of treatment.

Standardized monitoring of symptoms with multiple possible etiologies

Some but not all BMT centers reduce the stage in a target organ when another etiology of the symptom is documented (eg, infectious enteritis). Centers often do not publicly acknowledge this practice, introducing a major source of variation in multicenter studies. Multiple recorded etiologies for a single symptom (eg, GVHD, infection, and/or

conditioning regimen) create additional sources of error, especially when implausible etiologies are provided, such as attributing new onset diarrhea 24 days post-BMT to the conditioning regimen. To overcome the resulting confusion we created a structure that assigns a confidence level to the GVHD diagnosis. First, we collect symptomatic data weekly regardless of suspected etiologies, and etiologies in the differential diagnosis are recorded (eg, conditioning regimen, infection, GVHD, etc.). Symptoms are then categorized by facts and clinical action: GVHD **negative** (an alternative etiology is present and treated, GVHD is not under consideration); GVHD **possible** (GVHD is under consideration but GVHD treatment is not initiated); GVHD **probable** (treatment for GVHD was started but no biopsy performed); or **confirmed** (unequivocal pathologic evidence for GVHD). This system brings clarity, particularly to patients whose biopsies are non-diagnostic and whose symptoms were neither treated nor staged (Table 1). Data review is performed daily by the lead data coordinator, enabling queries for clarification (eg, were steroids started for GVHD or some other reason) while memories are still fresh. In our experience with the first 50 patients, queries decreased dramatically (ie, from ~70% to <10% requiring clarification) as the real-time review reinforced best data collection practices. Standardization of the confidence in diagnosis appears to reduce variation in what each center classifies as GVHD.

Early review and adjudication improves consistency

Clarification of missing or contradictory information is challenging when performed in temporal proximity to an event; it is virtually impossible when performed months to years later. Furthermore, data guidance is most consistently applied when entries are compared to source documents in real time and feedback regarding errors is immediate. An ambitious but attainable goal is to report 90% of GVHD staging within 2 weeks and 98% within 4 weeks. Central electronic review of data occurs daily, enabling queries for clarification or source documents to occur within 48 hours of data entry. Weekly contact with each site rapidly identifies and addresses data accuracy concerns. Best practices are shared with all data managers via email and monthly teleconferences (separate from the adjudication webinars described below), increasing consistency and cohesion among sites.

Even experienced data managers can misinterpret complex cases. However, clinicians with GVHD expertise rarely review data and then only for selected patients (eg, for a clinical trial audit). This widespread practice is a major source of unrecognized error even though the combination of standardized GVHD guidance and expert adjudication has been shown to assure GVHD data quality and reduce inter-center variability [1]. For example, the maximal GVHD grade reported *by the center* did not correlate with survival, whereas the maximal grade as adjudicated *by a central expert panel* did provide the expected correlation. The need for adjudication of maximal GVHD stage, when differences are clearest, underscores the heightened need for adjudication of GVHD stage at onset when the differences are significantly more subtle. We propose an adjudication process in order to improve data consistency for complicated cases. Cases for adjudication can be identified in three ways: First, computerized triggers flag unusual scenarios (eg, isolated liver GVHD, untreated GI GVHD, <50% skin rash treated with systemic steroids). These scenarios occur in approximately 10% of GVHD cases. Second, the data manager entering the data can flag a case for adjudication at the time of data entry (eg, the guidance does not cover a complex

clinical scenario). Third, the lead data coordinator can select cases based on potential discrepancies or error (eg, 3-day duration of systemic steroid treatment of GVHD). Preliminary experience indicates that approximately 15% of cases are flagged for possible adjudication. The lead data coordinator can query the local data manager for further clarification (which resolves approximately 50% of the cases). The remaining cases are then scheduled for adjudication by a central panel, short case summaries, and source documents redacted of patient identifiers distributed to the central panel. At least one physician and one data manager from each center can participate and communicate nuances to others at the site. Minutes of each adjudication webinar are circulated, and the GVHD guidance manual is updated appropriately. We expect this type of adjudication to reduce the time to implement new guidance and the number of times similar scenarios need to be discussed. As an example, pathology reports can be difficult to interpret when pathologic findings do not establish a single diagnosis. We created guidance (Table 1) during adjudication webinars that specify how to use biopsies to support a diagnosis of GVHD based on the biopsy report, therapeutic decisions, and differential diagnosis. This guidance is particularly useful in cases where the pathology is equivocal or non-diagnostic.

Conclusion

Inconsistent staging and grading of acute GVHD among centers makes comparisons unreliable and has likely contributed to the failure to replicate promising single center results in multicenter trials [9]. We have identified major sources of inconsistencies and have developed standardized GVHD guidance to reduce them. One benefit of this new web-based data collection system is its creation of confidence levels to indicate the level of diagnostic certainty. In the future, the degree of confidence can be incorporated into analyses relevant to the development of diagnostic laboratory tests. We developed this new data collection system using actual cases from multiple centers in order to address real world issues and facilitate its widespread adoption, which could improve the quality of clinical trials in acute GVHD.

References

- [1]. Weisdorf DJ, Hurd D, Carter S, Howe C, Jensen LA, Wagner J, et al. Prospective grading of graft-versus-host disease after unrelated donor marrow transplantation: a grading algorithm versus blinded expert panel review. *Biol Blood Marrow Transplant*. 2003; 9:512–518. [PubMed: 12931120]
- [2]. Jagasia M, Arora M, Flowers ME, Chao NJ, McCarthy PL, Cutler CS, et al. Risk factors for acute GVHD and survival after hematopoietic cell transplantation. *Blood*. 2012; 119:296–307. [PubMed: 22010102]
- [3]. Martin PJ, McDonald GB, Sanders JE, Anasetti C, Appelbaum FR, Deeg HJ, et al. Increasingly frequent diagnosis of acute gastrointestinal graft-versus-host disease after allogeneic hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2004; 10:320–327. [PubMed: 15111931]
- [4]. Perkins J, Field T, Kim J, Kharfan-Dabaja MA, Fernandez H, Ayala E, et al. A randomized phase II trial comparing tacrolimus and mycophenolate mofetil to tacrolimus and methotrexate for acute graft-versus-host disease prophylaxis. *Biol Blood Marrow Transplant*. 2010; 16:937–947. [PubMed: 20102746]

- [5]. Atkinson K, Horowitz MM, Biggs JC, Gale RP, Rimm AA, Bortin MM. The clinical diagnosis of acute graft-versus-host disease: a diversity of views amongst marrow transplant centers. *Bone Marrow Transplant.* 1988; 3:5–10. [PubMed: 3048470]
- [6]. Carnevale-Schianca F, Leisenring W, Martin PJ, Furlong T, Schoch G, Anasetti C, et al. Longitudinal assessment of morbidity and acute graft-versus-host disease after allogeneic hematopoietic cell transplantation: retrospective analysis of a multicenter phase III study. *Biol Blood Marrow Transplant.* 2009; 15:749–756. [PubMed: 19450760]
- [7]. Przepiorka D, Weisdorf D, Martin P, Klingemann HG, Beatty P, Hows J, et al. Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant.* 1994; 1995; 15:825–828. [PubMed: 7581076]
- [8]. Rowlings PA, Przepiorka D, Klein JP, Gale RP, Passweg JR, Henslee-Downey PJ, et al. IBMTR Severity Index for grading acute graft-versus-host disease: retrospective comparison with Glucksberg grade. *Br J Haematol.* 1997; 97:855–864. [PubMed: 9217189]
- [9]. Alousi AM, Weisdorf DJ, Logan BR, Bolanos-Meade J, Carter S, Difrizzo N, et al. Etanercept, mycophenolate, denileukin, or pentostatin plus corticosteroids for acute graft-versus-host disease: a randomized phase 2 trial from the Blood and Marrow Transplant Clinical Trials Network. *Blood.* 2009; 114:511–517. [PubMed: 19443659]

Table 1

Biopsy Results and Confidence Levels.

Pathology Results	Treatment Given for GVHD	Not Treated but GVHD Favored in Differential Diagnosis	Not Treated and GVHD Not Favored in Differential Diagnosis
Positive	Confirmed	Confirmed	Confirmed
Equivocal	Probable	Possible	Possible
Non-Diagnostic	Probable	Possible	Negative
Non-GVHD Etiology	Negative	Negative	Negative