VOLUME 46 · NUMBER 3 · September 2011

KJH

THE KOREAN JOURNAL OF HEMATOLOGY

REVIEW ARTICLE

Treatment of chronic graft-versus-host disease: Past, present and future

Paul J. Martin[1,2], Yoshihiro Inamoto[1,2], Paul A. Carpenter[1,3], Stephanie J. Lee[1,2], Mary E.D. Flowers[1,2]

[1]Division of Clinical Research, Fred Hutchinson Cancer Research Center, Departments of [2]Medicine and [3]Pediatrics, University of Washington, Seattle, WA, USA

p-ISSN 1738-7949 / e-ISSN 2092-9129
http://dx.doi.org/10.5045/kjh.2011.46.3.153
Korean J Hematol 2011;46:153-63.

Received on August 12, 2011
Accepted on August 19, 2011

*This study was supported by research grant CA98906 from the National Institutes of Health, Department of Health and Human Services. Y.I. is a recipient of the JSPS Postdoctoral Fellowships for Research Abroad.

Correspondence to
Paul J. Martin, M.D.
Fred Hutchinson Cancer Research Center, P.O. Box 19024, Seattle, WA 98109-1024, USA
Tel: +1-206-667-4798
Fax: +1-206-667-5255
E-mail: pmartin@fhcrc.org

Chronic GVHD was recognized as a complication of allogeneic hematopoietic cell transplantation more than 30 years ago, but progress has been slowed by the limited insight into the pathogenesis of the disease and the mechanisms that lead to development of immunological tolerance. Only 6 randomized phase III treatment studies have been reported. Results of retrospective studies and prospective phase II clinical trials suggested overall benefit from treatment with mycophenolate mofetil or thalidomide, but these results were not substantiated by phase III studies of initial systemic treatment for chronic GVHD. A comprehensive review of published reports showed numerous deficiencies in studies of secondary treatment for chronic GVHD. Fewer than 10% of reports documented an effort to minimize patient selection bias, used a consistent treatment regimen, or tested a formal statistical hypothesis that was based on a contemporaneous or historical benchmark. In order to enable valid comparison of the results from different studies, eligibility criteria, definitions of individual organ and overall response, and time of assessment should be standardized. Improved treatments are more likely to emerge if reviewers and journal editors hold authors to higher standards in evaluating manuscripts for publication.

Key Words Chronic graft-versus-host disease, Treatment, Phase II clinical trials, Review

INTRODUCTION - THE PAST

Allogeneic hematopoietic cell transplantation (HCT) is frequently complicated by acute and chronic graft-versus-host disease (GVHD) [1, 2]. Although considerable progress has been made in the development of methods to prevent or treat acute GVHD, similar progress in chronic GVHD has languished by comparison after the clinical and pathologic features of this syndrome were first described in 1980 [3]. Interest in this debilitating complication of HCT was rejuvenated when recommendations of the National Institutes of Health Consensus Conference on Criteria for Clinical Trials in Chronic Graft-versus-host disease were published in 2005-2006 [4-9].

The Consensus Conference recognized two major categories of GVHD, each with 2 subcategories [4]. Acute GVHD with onset before day 100 was defined as "classic GVHD." A separate category was recognized for persistent, recurrent or late-onset acute GVHD beyond day 100 after HCT. Chronic GVHD was separated from acute GVHD not by time from HCT but by the presence of diagnostic criteria or by distinctive findings supported by biopsy or other procedures. Classic chronic GVHD was defined by unambiguous chronic GVHD manifestations in the absence of abnormalities such as cutaneous erythema, liver function abnormalities, or gastrointestinal manifestations typical of acute GVHD. "Overlap syndrome" is a subcategory of chronic GVHD characterized by chronic GVHD in the presence of one or more manifestations typical of acute GVHD.

Chronic GVHD is a pleomorphic syndrome with "autoimmune" features that sometimes resemble clinical findings in scleroderma and Sjögren syndrome. The onset usually occurs between 3 and 15 months after HCT [10-13]. Risk factors associated with an increased risk of chronic GVHD include the use of a mobilized blood cell graft or an HLA-mismatched or unrelated donor, older patient age and a history of acute GVHD [12]. The risk of chronic GVHD

can be decreased by exhaustive depletion of T cells from the graft or by treatment of the recipient with rabbit antibodies specific for human T cells as part of the conditioning regimen before HCT [12, 14, 15]. Without these measures, approximately 30% to 50% of HCT recipients develop chronic GVHD [2, 16].

Chronic GVHD can affect multiple organs and sites, including the skin and subcutaneous connective tissues, lacrimal and salivary glands, oral mucosa, lungs, esophagus, joints, gastrointestinal tract and liver. The disease is characterized by immune dysfunction with an increased risk of infections and a 30% to 50% risk of mortality during the first 5 years after diagnosis [10, 11]. Chronic GVHD has been associated with a reduced risk of recurrent malignancy after HCT [17-22], but despite this benefit, survival is not improved [22]. A prognostic scoring system has recently been proposed based on factors present at the time of chronic GVHD diagnosis [13].

PRINCIPLES OF TREATMENT - THE PRESENT

To date, only 6 randomized phase III studies have ever been reported for initial treatment of chronic GVHD [23-28]. The study by Koc et al. [26] was the only one that indicated benefit. Results of this study suggested that treatment with cyclosporine reduced the amount of glucocorticoid treatment needed to control the disease, as indicated by a decreased frequency of avascular necrosis. The generally recommended approach for treatment of chronic GVHD involves continued administration of the calcineurin inhibitor used for GVHD prophylaxis together with prednisone initially at 1 mg/kg/day [2, 29, 30]. Strategies for tapering the dose of prednisone vary considerably, but as a general principle, efforts should be made to use the minimum dose that is sufficient to control GVHD manifestations. At our center, the dose of prednisone is tapered to an alternate-day schedule of administration after initial clinical improvement, which generally occurs within 6 weeks after starting treatment. The dose of prednisone is then tapered to 0.5 mg/kg every other day and generally continued until reversible manifestations of the disease resolve. The dose of prednisone is then gradually tapered with careful monitoring for recurrent manifestations of chronic GVHD. Doses of the calcineurin inhibitor are gradually decreased after treatment with prednisone has been withdrawn.

The median duration of treatment for chronic GVHD is approximately 2 years in patients who had HCT with marrow cells and approximately 3.5 years in those who had HCT with growth factor-mobilized blood cells [31]. The current therapeutic approach functions primarily to prevent immune-mediated damage, while awaiting the development of tolerance. Evidence to suggest that current treatments accelerate the development of immunological tolerance is lacking. The mechanisms that facilitate development of tolerance have not been defined.

Administration of medications to prevent infection with Pneumocystis jirovecii and encapsulated bacteria is necessary during treatment for chronic GVHD [32]. Some patients may need topical or systemic treatment to prevent mucocutaneous candida infection. Patients at risk of Varicella zoster activation should be given antiviral prophylaxis, and CMV monitoring and preemptive treatment is necessary in patients at risk of CMV infection [33]. Activation of CMV during the first 3 months after HCT suggests an increased risk of subsequent reactivation in patients with chronic GVHD. Systemic immunosuppressive treatment should be administered at the lowest effective dose in order to minimize the risk of infections and other complications. Many steroid-related complications can be avoided or at least minimized by an alternate-day schedule of administration [34], and topical treatment can be used to minimize the need for systemic treatment [35]. Bone mineral density should be monitored yearly, and losses should be minimized through weight bearing exercise, administration of calcium and vitamin D supplements and hormone replacement.

Indications for secondary treatment include worsening manifestations in a previously affected organ, development of manifestations in a previously unaffected organ, absence of improvement after 1 month of treatment, or inability to decrease the dose of prednisone below 1.0 mg/kg/day within 2 months [30, 36]. Numerous clinical trials have been carried out to evaluate approaches for secondary treatment of chronic GVHD. To date, no consensus has been reached regarding the optimal choice of agents for secondary treatment, and clinical management is generally approached through empirical trial and error [36]. Treatment choices are based on physician experience, ease of use, need for monitoring, risk of toxicity and potential exacerbation of pre-existing co-morbidity.

MYCOPHENOLATE MOFETIL: A CASE STUDY ILLUSTRATING CURRENT PROBLEMS

Progress in the clinical management of chronic GVHD has been slowed by limited insight into the pathogenesis of the disease and the mechanisms that lead to development of immunological tolerance. In the absence of pathophysiologic understanding, physicians must rely on personal or published empirical experience in making decisions regarding treatment. In principle, results of treatment in patients with “steroid-refractory” or “steroid-resistant” chronic GVHD could be used to identify promising agents for initial treatment. Effective agents would be expected to decrease reliance on glucocorticoids and could conceivably decrease the duration of time needed for resolution of the disease.

A variety of retrospective and phase II studies suggested that MMF could be used successfully for secondary treatment of chronic GVHD. In results of a survey published in 2002, nearly 80% of clinicians reported that they had used mycophenolate mofetil (MMF) with great success or at least some success for treatment of chronic GVHD [37]. In another survey proposing a hypothetical scenario describing a case of high-risk chronic GVHD, 54% of the respondents indicated that they

would add MMF for initial treatment of chronic GVHD [38]. These results supported a formal test of the hypothesis that the addition of MMF to standard initial treatment could improve outcomes for patients with chronic GVHD.

We therefore conducted a prospective, double-blind, randomized phase III clinical trial to test this hypothesis [27]. The primary endpoint was resolution of reversible manifestations of chronic GVHD within 2 years after enrollment, before death or the onset of recurrent malignancy and without the need for secondary systemic treatment. It was expected that the use of MMF would shorten the time to response, decrease systemic steroid exposure, and decrease the risk of transplant-related mortality without increasing the risk of recurrent malignancy, thereby potentially improving overall survival. Results of the trial, however, did not show any benefits of treatment with MMF. Potential reasons for the negative results were thoroughly explored. The absence of success in this randomized trial could not be attributed to an imbalance of risk factors between the arms, sub-optimal dosing of MMF or non-adherence with administration of the study drug. Hence, this clinical trial definitively demonstrated that addition of MMF to standard initial treatment did not improve outcomes for patients with chronic GVHD.

These unexpected results conflicted with previously prevailing clinical impressions and motivated a careful review of prior reports evaluating the use of MMF for treatment of chronic GVHD. Overall results of 9 such studies suggested that secondary treatment with MMF produced a 20% complete response rate and a 65% complete or partial response rate (Table 1) [39-47]. One of these studies also evaluated results in 10 patients who received MMF as part of the initial treatment regimen for chronic GVHD [45]. Seven of the 10 patients had a complete response, and 2 had a partial response, yielding an overall 90% rate of complete or partial response. In addition, a further study from our center had shown that the proportion of patients who discontinued systemic immunosuppressive treatment after resolution of reversible abnormalities increased progressively from 9% to 17% and 26% at 1, 2 and 3 years, respectively, after starting treatment with MMF [48].

Similar discrepancies have been observed in studies to evaluate the efficacy of thalidomide for treatment of chronic GVHD. Results of 6 retrospective studies and prospective phase II clinical trials suggested favorable outcomes with the use of thalidomide for secondary treatment of chronic GVHD [49-54]. The two randomized prospective studies testing the use of thalidomide for primary treatment of chronic GVHD, however, showed no benefit [24, 25].

Results of the randomized trials defied expectations coming from at least 16 studies evaluating the use of MMF or thalidomide for treatment of chronic GVHD. At least 2 explanations could be proposed to explain this discrepancy. 1) Results of secondary treatment might not predict efficacy as an added agent for primary treatment, perhaps because most patients do not need additional agents in order to gain maximal benefit from initial treatment. Experience at our center, however, has indicated that systemic treatment is changed in approximately 60% of patients during the first 3 years because of inadequate response to primary therapy for chronic GVHD [55]. 2) Alternatively, previous studies might have had unrecognized limitations leading to overstated expectations.

Table 1. Response rates in prior studies of mycophenolate mofetil.

Study	Type	N	CR	PR	CR+PR (%)
1	Retrospective	26	2	10	12 (46)
2	Prospective	5	2	3	5 (100)
3	Prospective	21	5	8	13 (62)
4	Prospective	15	2	7	9 (60)
5	Retrospective	11	3	5	8 (73)
6	Retrospective	13	1	9	10 (77)
7	Retrospective	13	4	5	9 (69)
8	Retrospective	24	5	13	18 (75)
9	Prospective	11	4	3	7 (64)
	Total	139	28 (20%)	63	91 (65)

PROPOSED QUALITY INDICATORS FOR EVALUATING REPORTS

Previous publications have identified quality indicators for evaluation of phase III clinical trials [56, 57], but to our knowledge, similar quality criteria have not been previously proposed for phase II trials and retrospective studies. Therefore, before embarking on a detailed review of the 10 previous studies evaluating the use of MMF, we developed a list of 10 quality indicators that could be used to characterize an ideal prospective phase II clinical trial or retrospective study of treatment for GVHD. The proposed quality indicators are summarized below.

1. Adequately defined eligibility criteria

Inclusion and exclusion criteria should specify affected sites, severity of manifestations, and prior treatment used to define the cohort. Exclusion criteria should indicate whether factors such as the presence of infection, inability to tolerate the study treatment, presence of persistent malignancy or low performance score were used to define the cohort. Studies intended to evaluate treatment of "steroid-refractory" GVHD should indicate the glucocorticoid dose and duration of treatment used to define the cohort. Eligibility criteria are typically more precisely defined for prospective studies than for retrospective studies. Data from retrospective studies describing all patients who received the study treatment of interest are difficult to interpret unless additional selection criteria are applied to improve homogeneity within the study cohort.

2. Documented minimization of bias in the selection of patients

Readers should be given enough information to determine

whether the characteristics of the patients included in a study are representative of the more general population of patients with chronic GVHD. Risk factors that could affect outcome should be delineated. Ideally, either an historical or contemporaneous cohort should be identified for comparison, and any differences in the prevalence of risk factors between the study cohort and the comparison cohort should be noted. The use of randomization to define cohorts helps to ensure the absence of bias, but this procedure does not ensure that the study cohort is representative of the more general population of patients with the indication of interest. Enrollment of all consecutive patients who meet eligibility criteria can ensure that the cohort is representative of the more general population, but this approach would raise concerns about the adequacy of informed consent. Thus, comparisons of demographics and risk factors between patients who participated and those who did not are crucial.

3. Consistent treatment regimen

The study treatment of interest should be administered in a consistent manner in dose, schedule and duration of administration. Differences in dose, schedule or duration of administration can be addressed by stratified analysis of each specific subgroup. As much as possible, concomitant treatment with immunosuppressive agents other than glucocorticoids should also be administered in a consistent manner in order to facilitate the interpretation of results. Such consistency greatly improves the ability to interpret results and to confirm the results in subsequent studies. Concomitant treatment can be standardized more easily in studies of initial therapy for standard or high-risk disease and for secondary therapy than in studies of subsequent therapies. For third-line or subsequent therapy, such consistency is feasible only if prior treatment with agents other than glucocorticoids is discontinued.

4. Objective criteria for organ response

Categorical criteria should be defined for complete response, partial response, no change, and worsening for each organ or site affected by GVHD, even if organ response criteria have not been validated, since conclusions of the study are based on response rates. Definitions require formal assessment at baseline and at the comparison time point. In many studies, partial response was defined as "at least 50% improvement" in disease manifestations. This terse, and likely oversimplified, definition meets the formal criterion of objectivity but suggests that the response assessment actually reflects a general overall impression, as opposed to a detailed comparison of changes in chronic GVHD manifestations in each organ between baseline and the assessment time.

5. Unambiguous criteria for overall response

The definition of overall response is distinct from the criteria for organ response. Overall responses are often defined according to the overall pattern of organ responses. At a minimum, overall partial response indicates improvement in at least one organ. The category assigned for patients with improvement in one organ but deterioration in another organ should be clearly stated.

6. Specified time for assessment of response

To facilitate comparisons between studies, at least one specified time point should be used for assessment of response, and the data for this assessment should be shown. Additional information can be shown as a time to event analysis. The number of patients who died or had recurrent malignancy before the assessment time point should be specified, and results should clearly indicate whether these patients were excluded from consideration in the assessment of response or whether they were included as non-responders. Tabulation of results according to "best response" or "last value carried forward" is not appropriate, since these categories do not reflect clinical benefit at a specific time point.

7. Concomitant treatment taken into account

New systemic treatment for GVHD added after enrollment but before the assessment time point because of inadequately controlled disease manifestations should be categorized as non-response. Even in studies that use "best response" as the endpoint, the text should state whether response was evaluated before any new systemic treatment was added. Changes in glucocorticoid dose should be described, but a temporary small increase in glucocorticoid dose during a taper should not be categorized as non-response, because temporary flares of GVHD activity cannot be avoided when conscientious efforts are made to determine the minimum glucocorticoid dose needed to control GVHD.

8. Well-established control benchmark

A specific historical or concurrent control benchmark should be used to establish a null hypothesis for the primary endpoint. Response criteria for the benchmark and study cohorts should be identical or closely similar.

9. Statistical hypothesis and power estimate

The methods should provide values for the null and alternative hypotheses and for the one-sided or two-sided type 1 error, together with estimates of statistical power and the necessary sample size. Although these considerations might be difficult to apply in retrospective studies, they should always be applied in prospective studies.

10. Survival

The results should show survival of the cohort from the onset of study treatment. Kaplan-Meier curves should show tic marks depicting end of follow-up, especially if the minimum follow-up time for surviving patients is less than 6 months. Alternatively, results can be shown in tables indicating time to death or last follow-up for each patient. When response definitions differ, survival data provide the only gauge that can be used as a simple and universally applicable method for comparisons with other studies.

Table 2. Quality of prior reports of studies testing mycophenolate mofetil.

Quality criterion	Study number										Total
	1	2	3	4	5	6	7	8	9	10	
Eligibility criteria		Y		Y						Y	3
Minimization of selection bias											0
Consistent treatment regimen	Y				Y	Y					3
Objective response criteria			Y	Y							2
Overall response criteria				Y						Y	2
Time of assessment										Y	1
Concomitant treatment		Y	Y	Y	Y		Y		Y	Y	7
Historical benchmark											0
Statistical hypothesis											0
Survival curve						Y			Y		2
Total	1	2	2	4	2	2	1	0	2	4	

Table 3. Initial agreement between evaluators.[a)]

Quality criterion	Agreement
Eligibility criteria	87%
Minimization of selection bias	97%
Consistent treatment regimen	92%
Objective response criteria	75%
Overall response criteria	78%
Time of assessment	88%
Concomitant treatment	72%
Historical benchmark	97%
Statistical hypothesis	98%
Survival curve	87%

[a)]Each of the 60 selected reports was independently evaluated by 2 reviewers. Results in the table indicate the percent agreement between the 2 reviewers for each quality criterion.

EVALUATION OF PRIOR REPORTS FOR STUDIES OF MMF

Two individuals (YI and PM) independently reviewed the 10 prior reports of studies testing the use of MMF for secondary treatment of chronic GVHD [39-48]. Reports were evaluated according to whether each quality criterion was met or not, based on careful reading of the text. Differences in scores were reconciled by joint review to arrive at a consensus. Since the purpose of publication is to persuade others, application of the criteria was very strict, and no credit was given if the text did not address the criterion or if the text was not clear. Therefore, in many cases, deficiencies in the report might not have been representative of a study as it was actually conducted.

Results for the 10 studies of MMF are summarized in Table 2. Scores at the bottom of the table represent the total number of criteria met by each report. One report failed to meet any of the 10 criteria. Two reports met 4 criteria, and none had higher scores. The mean score for the 10 reported studies was 2.0. Scores at the right margin of the table represent the number of reports that met each criterion. None of the reports attempted to demonstrate that bias had been minimized in the selection of patients, used an historical or contemporaneous benchmark or tested a statistical hypothesis. Only one report had a specified time of assessment, and only two had objective response criteria and well-defined overall response criteria. Three reports employed a consistent treatment regimen, while 7 accounted for possible effects of concomitant treatment.

COMPREHENSIVE REVIEW OF PRIOR REPORTS

Results of the review of reports for studies testing MMF prompted a more comprehensive review of studies testing systemic agents for secondary treatment of chronic GVHD published between 1990 and 2011. We searched the Medline (PubMed) database using a broad search strategy to identify studies evaluating secondary treatment of chronic GVHD. The search was conducted using the terms "Chronic graft versus host disease" and "Treatment" excluding '"Review." Relevant references in the publications identified were also reviewed. Both retrospective and prospective studies were included, but studies with cohorts containing fewer than 10 patients (N=26), phase III studies and case reports were excluded. A total of 60 studies were selected for review [39-54, 58-101]. Initial agreement between the two reviewers was high, ranging between 72% and 98% (Table 3).

Across the 60 studies, 17 different agents were evaluated (Fig. 1). Extracorporeal photopheresis was the most frequently studied agent (N=17) followed by mycophenolate mofetil (N=10), thalidomide (N=6), sirolimus or everolimus (N=4) and rituximab (N=4). The distribution of scores representing the total number of criteria met by each report ranged from a low of 0 (N=6) to 8 (N=1) [61] (Fig. 2). The mean score for all 60 reports was 2.5. The mean score for prospective studies (N=31) was 3.1, compared to 1.8 for retrospective studies (N=29). The mean score for multicenter studies (N=7) was 3.6, compared to 2.3 for single-center studies (N=53).

Approximately 35% to 45% of all reports provided adequate information regarding eligibility criteria, organ response criteria, overall response criteria, concomitant treatment and overall survival (Table 4). Only 22% of the reports had a specified time for assessment of response, and less than 10% of the reports documented an absence of bias in the selection of patients, used a consistent treatment regimen, or tested a formal statistical hypothesis on the basis of a benchmark from a contemporaneous or historical cohort. The percentage of reports fulfilling quality indicators was generally higher for prospective studies than for retrospective studies (Table 4).

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.