

A phase II/III randomized, multicenter trial of prednisone/sirolimus versus prednisone/sirolimus/calcineurin inhibitor for the treatment of chronic graft-versus-host disease: BMT CTN 0801

Paul A. Carpenter,¹ Brent R. Logan,² Stephanie J. Lee,¹ Daniel J. Weisdorf,³ Laura Johnston,⁴ Luciano J. Costa,⁵ Carrie L. Kitko,⁶ Javier Bolaños-Meade,⁷ Stefanie Sarantopoulos,⁸ Amin M. Alousi,⁹ Sunil Abhyankar,¹⁰ Edmund K. Waller,¹¹ Adam Mendizabal,¹² Jiaxi Zhu,¹² Kelly A. O'Brien,¹² Aleksandr Lazaryan,³ Juan Wu,¹² Eneida R. Nemecek,¹³ Steven Z. Pavletic,¹⁴ Corey S. Cutler,¹⁵ Mary M. Horowitz² and Mukta Arora³ on behalf of the BMT CTN.

¹Fred Hutchinson Cancer Research Center, Seattle, WA; ²Medical College of Wisconsin, Milwaukee, WI; ³University of Minnesota, Minneapolis, MN; ⁴Stanford Hospital and Clinics, Stanford, CA; ⁵University of Alabama at Birmingham, AL; ⁶Vanderbilt University School of Medicine, Nashville, TN; ⁷Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD; ⁸Duke University Medical Center, Durham, NC; ⁹University of Texas MD Anderson Cancer Center, Houston, TX; ¹⁰The University of Kansas Medical Center, Kansas City, KS; ¹¹Emory University School of Medicine, Atlanta, GA; ¹²The Emmes Corporation, Rockville, MD; ¹³Oregon Health and Sciences University, Portland, OR; ¹⁴Experimental Transplantation and Immunology Branch, National Cancer Institute, Bethesda, MD and ¹⁵Dana Farber Cancer Institute, Boston, MA, USA

ABSTRACT

Initial therapy of chronic graft-versus-host disease is prednisone ± a calcineurin-inhibitor, but most patients respond inadequately. In a randomized, adaptive, phase II/III, multicenter trial we studied whether prednisone/sirolimus or prednisone/sirolimus/photopheresis was more effective than prednisone/sirolimus/calcineurin-inhibitor for treating chronic graft-versus-host disease in treatment-naïve or early inadequate responders. Primary endpoints of this study were proportions of subjects alive without relapse or secondary therapy with 6-month complete or partial response in phase II, or with 2-year complete response in phase III. The prednisone/sirolimus/photopheresis arm closed prematurely because of slow accrual and the remaining two-drug versus three-drug study ended in phase II due to statistical futility with 138 evaluable subjects. The two-drug and three-drug arms did not differ in rates of 6-month complete or partial response (48.6% versus 50.0%, $P=0.87$), or 2-year complete response (14.7% versus 15.5%, $P=0.90$). Serum creatinine values >1.5 times baseline were less frequent in the calcineurin-inhibitor-free arm at 2 months (1.5% versus 11.7%, $P=0.025$) and 6 months (7.8% versus 24.0%, $P=0.016$). Higher adjusted Short Form-36 Physical Component Summary and Physical Functioning scores were seen in the two-drug arm at both 2 months ($P=0.02$ and $P=0.04$, respectively) and 6 months ($P=0.007$ and $P=0.001$, respectively). Failure-free survival and overall survival rates at 2 years were similar for patients in the the two-drug and three-drug arms (48.6% versus 46.2%, $P=0.78$; 81.5% versus 74%, $P=0.28$). Based on similar long-term outcomes, prednisone/sirolimus is a therapeutic alternative to prednisone/sirolimus/calcineurin-inhibitor for chronic graft-versus-host disease, being easier to administer and better tolerated. *Clinicaltrials.gov identifier: NCT01106833.*



Haematologica 2018
Volume 103(11):1915-1924

Correspondence:

pcarpent@fredhutch.org

Received: April 7, 2018.

Accepted: June 26, 2018.

Pre-published: June 28, 2018.

doi:10.3324/haematol.2018.195123

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/103/11/1915

©2018 Ferrata Storti Foundation

Material published in Haematologica is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>. Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions: <https://creativecommons.org/licenses/by-nc/4.0/legalcode>, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



Introduction

Chronic graft-versus-host disease (GvHD) is the major treatment-related complication among patients who survive after allogeneic hematopoietic cell transplantation. The protracted duration of chronic GvHD and its protean and sometimes irreversible organ manifestations makes it the leading cause of impaired immunity, compromised functional status, and late treatment-related deaths.^{1,2} Standard immunosuppressive therapy with prednisone \pm a calcineurin inhibitor (CNI) has not changed for three decades, but most patients respond inadequately with slow and most often incomplete control of their disease. The results of six phase III chronic GvHD trials that tested prednisone or prednisone/CNI backbones with or without an experimental immunosuppressive therapy have not changed clinical practice, because the treatments tested in the experimental arms did not improve outcomes due to lack of efficacy and/or more toxicity.³⁻⁸ Between 2005-2007, reports of phase II trials from single centers showed promising results for prednisone combined with sirolimus, rituximab, mycophenolate mofetil, pentostatin or extracorporeal photopheresis (ECP).⁹⁻¹³ At the same time, renewed understanding of regulatory T cells (Treg) led to the hypothesis that experimental therapies permissive of Treg expansion would abrogate GvHD better than CNI-containing (control) immunosuppressive therapy. Sirolimus and ECP are permissive of Treg expansion.¹⁴⁻¹⁸ Because of their acceptance in clinical practice sirolimus and ECP were considered good candidates to add to prednisone in clinical trials investigating chronic GvHD.

This background, together with the 2005 National Institutes of Health (NIH) chronic GvHD criteria,¹⁹ plus other efforts designed to propel the field forward,^{20,21} motivated the Blood and Marrow Transplant Clinical Trials Network (BMT CTN) to conduct an intervention trial for chronic GvHD. The intent was an adaptive phase II/III design to minimize between-phase downtime. The primary purpose of phase II was to select the more promising of two CNI-free approaches (prednisone/sirolimus and prednisone/sirolimus/ECP) in order to proceed seamlessly into a definitive phase III study against a CNI-containing comparator arm (prednisone/sirolimus/CNI). The study also provided an opportunity to use the NIH diagnostic and response criteria prospectively and to validate them.

Methods

Patients

Adult and pediatric allogeneic hematopoietic cell transplant recipients were eligible if they had classic chronic GvHD \pm acute GvHD (overlap subtype) that met 2005 NIH diagnostic consensus criteria.¹⁹ Eligibility criteria were broad and allowed a period of steroid exposure prior to enrollment to ensure congruence with standard practice. Thus, eligible patients were either: newly diagnosed patients, defined as individuals who had received <14 days of prednisone (or equivalent) before randomization to the study therapy, or previously treated, but responding inadequately after ≤ 16 weeks of initial therapy with prednisone and/or a CNI \pm an additional non-sirolimus agent started at the time that the chronic GvHD was diagnosed. Major reasons for exclusion were patients with late persistent acute GvHD or recurrent acute GvHD only, patients unable to begin prednisone at a dose of 0.5 mg/kg day (or equivalent), patients already receiving sirolimus

for treatment of chronic GvHD, and patients already receiving sirolimus (for prophylaxis or treatment of acute GvHD) along with prednisone at ≥ 0.25 mg/kg/day (or equivalent) \pm additional agents. Patients were also ineligible if they had an invasive fungal or viral infection not responding to appropriate therapies, a creatinine clearance <50 mL/min/1.73 m² based on the Cockcroft-Gault (adults) or Schwartz (age ≤ 12 years) formula, an absolute neutrophil count $<1.5 \times 10^9/L$, a requirement for platelet transfusion, or a progressive or recurrent malignancy defined other than by quantitative molecular assays. Institutional review boards at all participating centers provided ethics approval. All patients or their parents signed informed consent to participation in the trial in accordance with the Declaration of Helsinki.

Treatment plan

The starting dose of prednisone (or prednisone-equivalent) was 1 mg/kg once daily, unless contraindicated, in which case the prednisone dose began at 0.5-1 mg/kg once daily. It was recommended that the dose of prednisone was tapered down, over 4-8 weeks to reach a dose of 0.5-1 mg/kg every other day, with the tapering starting within 2 weeks after the first evidence of GvHD improvement. Once an every-other-day prednisone (or equivalent) regimen was achieved, this dose remained constant for 10-12 weeks until all reversible chronic GvHD manifestations resolved. A second taper was then attempted and could follow individual institutional guidelines, but it was recommended that the extent of the tapering be approximately calibrated to the magnitude of an individual patient's every-other-day prednisone dose. CNI therapy was continued by targeting trough serum levels of 5-10 ng/mL for tacrolimus (120-200 ng/mL for cyclosporine). The sirolimus therapy began at a dose of 2 mg orally once daily (1 mg/m² per day if the patient weighed <40 kg) to target trough serum levels of 3-12 ng/mL. Supportive care was provided in accordance with institutional guidelines reflecting standard practices appropriate for chronic GvHD.²²

Study design

The trial was designed as an adaptive phase II/III randomized, open-label, prospective study of three treatments for chronic GvHD (Figure 1, *Online Supplementary Figures S1 and S2*). Phase II included two parallel, 100-patient, randomized trials, comparing prednisone/sirolimus or prednisone/sirolimus/ECP *versus* identical (prednisone/sirolimus/CNI) comparator arms. The primary objective of phase II was to estimate the proportion of study subjects at 6 months after randomization with complete or partial response, who were alive without relapse or receipt of secondary immunosuppressive therapy. A sufficiently promising phase II result would determine whether the trial would proceed into phase III with additional accrual, would continue as phase II, but be followed for longer phase III endpoints without additional accrual, or would end in failure without further follow-up (*Online Supplementary Figure S2*). Full details of the study design, endpoint definitions, statistical analysis and study timeline are contained in the *Online Supplementary Material*.

Results

Patients

One hundred patients were evaluated for the phase II primary endpoint after all had completed 6 months of follow-up. The Z-statistic comparing complete/partial response rates (51% *versus* 50%, $Z=0.11$, stopping boundary $Z_c \leq 0.9$) did not support proceeding to phase III and, together with pre-specified outcome scenarios, guided the

Data and Safety Monitoring Board recommendation to suspend further phase III accrual; however, all 151 enrolled subjects were followed for phase III endpoints (*Online Supplementary Figure S2B*). The endpoint review committee response adjudication determined that 13 (10%) subjects had not satisfied the NIH criteria for the diagnosis of chronic GvHD at enrollment and were ineligible (*Online Supplementary Figure S1C*); a screening diagnostic checklist for NIH-defined chronic GvHD was required after July 24, 2013. One-hundred thirty-eight remaining subjects were randomized: 72 were assigned to two drugs and 66 to three drugs. Endpoint review committee-adjudicated complete/partial response rates and provider-reported complete/partial response rates were concordant at 6 months (Cohen $\kappa = 0.78$) and at 2 years ($\kappa = 0.88$). Participants lost to follow-up, or who relapsed, died, or began secondary immunosuppressive therapy were excluded from these agreement tests.

Patients' demographic and transplant characteristics were mostly similar between treatment groups (Table 1). The median age was 50.2 years in the two-drug arm and 54.7 years in the three-drug arm. Characteristics that were more frequent in the two-drug arm were male gender (66.7% *versus* 50%), Hispanic ethnicity (11.1% *versus* 0%), an underlying diagnosis of acute leukemia (65.3% *versus* 37.9%), transplants done for early stage disease (63.9% *versus* 40.9%), myeloablative conditioning (65.3% *versus* 43.9%), and CNI/methotrexate-based GvHD prophylaxis (61% *versus* 33.3%). Lymphoma was less frequent in the two-drug arm (6.9% *versus* 25.8%). Donor or stem cell sources did not differ between arms.

Chronic GvHD characteristics were similar between the two-drug and three-drug arms (Table 2), including the proportion of subjects with high-risk chronic GvHD (51.4% *versus* 48.5%), median time from hematopoietic cell transplant to enrollment (7.3 *versus* 7.7 months), and median time from chronic GvHD diagnosis to enrollment (7 *versus* 10 days). Sixty-two percent of the study subjects were enrolled within 14 days of a new diagnosis of chronic GvHD and 38% were responding inadequately to initial therapy; classic and overlap subtypes were equally represented. Global severity scores were mainly moderate (67%), then mild (21%) and severe (12%); 43% of the patients had progressive onset chronic GvHD. The organs involved were the skin (76%), mouth (74%), liver (71%), eyes (54%), gastro-intestinal tract (46%), joints/fascia (24%), genital tract (17%), and lungs (15%) with the pattern of involvement being similar between the treatment groups except for genital tract involvement, which was more common in the group treated with three drugs (10% *versus* 26%, but only 63% of the study population were scored).

Response and survival

The treatment success rate at 6 months was 48.6% [95% confidence interval (CI): 36.7%-60.7%] with two drugs *versus* 50.0% (95% CI: 37.4%-62.6%) with three drugs ($P=0.87$). The rate of more stringently defined treatment success at 2 years (complete responses only) was 14.7% (95% CI: 7.3%-25.4%) with two drugs *versus* 15.5% (95% CI: 7.4%-27.4%) with three drugs ($P=0.90$) (Table 3). Since very good partial response is clinically rel-

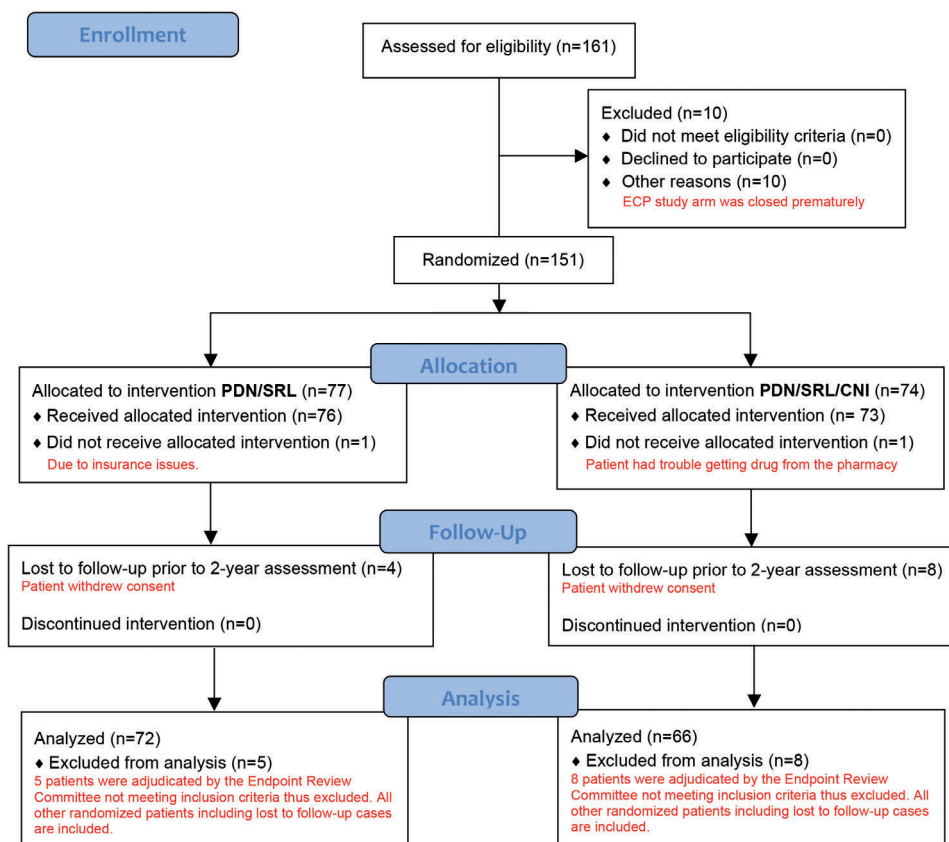


Figure 1. BMT CTN 0801 Consort flow diagram. ECP: extracorporeal photopheresis; PDN: prednisone; SRL: sirolimus; CNI: calcineurin inhibitor.

evant, we compared the proportion of patients with complete response/very good partial response who otherwise satisfied the 2-year treatment success definition. By this measure, the 29.4% (95% CI: 19.0-41.7%) in the two-drug arm did not differ from the 22.4% (95% CI: 12.5-35.3%) in the three-drug arm ($P=0.37$).

To address whether certain categories of chronic GvHD responded differently in the study arms overall, in univariate analysis, we compared subjects with mild, moderate and severe chronic GvHD, and *de novo*, quiescent onset and progressive chronic GvHD. There were no statistically significant differences between these groups in terms of the odds of treatment success at 6 months or 2 years. Similarly, no particular individual organ involvement was more likely to be associated with treatment success using either treatment regimen.

The overall survival rate at 2 years was 81.5% in the group treated with two drugs *versus* 74% in the group treated with three drugs ($P=0.28$) (Figure 2A). Progression-free survival rates were 78.6% with two drugs *versus* 67.3% with three drugs ($P=0.14$) (Figure 2B) and the 2-year failure-free survival rates were 48.6% with two drugs *versus* 46.2% with three drugs ($P=0.78$) (Figure 2C,D). The cumulative incidence of relapse (10.1% *versus* 14.9%,

$P=0.40$), non-relapse mortality (5.6% *versus* 11.1%, $P=0.26$), and the cumulative incidence of secondary immunosuppressive therapy (38.5% *versus* 29.4%, $P=0.22$) were not different between the groups treated with two or three drugs. By 2 years, 13 deaths had occurred in the two-drug arm and 16 in the three-drug arm (Online Supplementary Table S1). Almost one-third of deaths were due to GvHD (30.8% in the prednisone/sirolimus group *versus* 31.3% in the prednisone/sirolimus/CNI group). Recurrent or progressive malignancy was the primary cause of death in 30.8% of the prednisone/sirolimus group *versus* 12.5% in the prednisone/sirolimus/CNI group ($P=0.36$). There were two deaths (15.4%) primarily from infection in the two-drug arm and three (18.8%) in the three-drug arm.

Curves depicting cumulative incidence of discontinuation of systemic immunosuppressive therapy by 2 years overlapped, with the values at 2 years being 23.2% in the two-drug arm and 20% in the three-drug arm ($P=0.71$) (Figure 3). Mean (standard deviation) daily glucocorticoid doses (mg/kg) at baseline were 0.9 (0.4) for the group given prednisone/sirolimus *versus* 0.8 (0.2) for the group given prednisone/sirolimus/CNI ($P=0.13$). From baseline to 1 year the mean (standard deviation) daily dose reduction was 0.7 mg/kg (0.3) for the prednisone/sirolimus group and 0.5 mg/kg (0.5) for the prednisone/sirolimus/CNI group.

Table 1. Demographic and hematopoietic cell transplant characteristics.

	Two drugs N = 72 (%)	Three drugs N = 66 (%)
Gender		
Female	24 (33.3)	33 (50.0)
Male	48 (66.7)	33 (50.0)
Ethnicity		
Hispanic or Latino	8 (11.1)	0 (0.0)
Not Hispanic or Latino	63 (87.5)	59 (89.4)
Unknown/not answered	1 (1.4)	7 (10.6)
Race		
White	54 (75.0)	57 (86.4)
Non-white	13 (18.1)	9 (13.6)
Unknown/not answered	5 (6.9)	0 (0.0)
Age, years		
Median	50.2	54.7
<20	5 (6.9)	2 (3.0)
≥20	67 (92.1)	64 (97.0)
Primary disease		
Acute leukemia	47 (65.3)	25 (37.9)
Chronic leukemia	7 (9.7)	8 (12.1)
MDS/MPS	12 (16.7)	9 (13.6)
Lymphoma	5 (6.9)	17 (25.8)
Other	1 (1.4)	7 (10.6)
Disease stage		
Early	46 (63.9)	27 (40.9)
Intermediate	11 (15.3)	21 (31.8)
High risk	15 (20.8)	18 (27.3)
Stem cell type		
Bone marrow (BM)	5 (6.9)	5 (7.6)
Peripheral blood stem cells (PBSC)	64 (88.9)	60 (90.9)
Cord blood	3 (4.2)	1 (1.5)
Donor source (for BM/PBSC)		
Related	29 (42.0)	38 (58.5)
Unrelated	40 (58.0)	27 (41.5)
Type of conditioning regimen		
Myeloablative	47 (65.3)	29 (43.9)
Non-myeloablative or reduced intensity	25 (34.7)	37 (56.1)

continued from the previous column

	Two drugs N = 72 (%)	Three drugs N = 66 (%)
GvHD prophylaxis after transplant		
CNI + methotrexate ± other	44 (61.0)	22 (33.3)
CNI + MMF	9 (12.5)	15 (22.7)
CNI + MMF ± Other	5 (6.9)	7 (10.6)
CNI + MMF + ATG	0 (0.0)	1 (1.5)
CNI + sirolimus	3 (4.2)	5 (7.6)
CNI + corticosteroids	2 (2.8)	0 (0.0)
CNI only	5 (6.9)	5 (7.6)
Other regimen	4 (5.6)	11 (16.7)
Donor age, years		
Mean (SD)	38.5 (12.9)	44.4 (15.2)
Median	38.8	47.0
Donor age, years		
<20	3 (4.8)	4 (6.6)
20-40	31 (50.0)	21 (34.4)
41-60	24 (38.7)	27 (44.3)
>60	4 (6.5)	9 (14.8)
Donor gender		
Female	35 (48.6)	26 (39.4)
Male	37 (51.4)	38 (57.6)
Unknown	0 (0.0)	2 (3.0)
Donor recipient gender mismatch		
Female-female	14 (19.4)	10 (15.6)
Female-male	21 (29.2)	16 (25.0)
Male-female	10 (13.9)	21 (32.8)
Male-male	27 (37.5)	17 (26.6)

Two drugs: prednisone/sirolimus, Three drug: prednisone/sirolimus/calcineurin-inhibitor; MDS/MPS: myelodysplastic syndrome/myeloproliferative syndrome; BM: bone marrow; PBSC: peripheral blood stem cells; CNI: calcineurin inhibitor; MMF: mycophenolate mofetil; ATG: anti-thymocyte globulin; GvHD: graft-*versus*-host disease; SD: standard deviation.

Biomarkers

Correlative biology studies were attempted using a pre-specified analysis of plasma B-cell activating factor (BAFF) levels determined by enzyme-linked immunosorbent assay, and CD3⁺CD4⁺CD35⁺CD127⁻ Treg and CD19⁺ B-cell enumeration in blood by flow cytometry. Data were available for 99% of patients at baseline, 86% at 60 days and 76% at 180 days after beginning study therapy. Median Treg levels (cells per microliter) in both arms were identical, 26, at baseline ($P=0.96$), and were lower, 18, at day 60 ($P=0.87$). At day 180 after starting study therapy, the median Treg number was higher in the group treated with two drugs, 21 (range, 0-68) than in the group treated with three drugs 14.5 (range, 1-239) ($P=0.04$), but there was no significant difference in Treg: conventional

T-cell (Tcon) ratios. Treatment success at day 180, irrespective of treatment arm, was associated with significantly higher median Treg levels at day 60, 21 (range, 2-78) versus 10 (range, 1-53) among those with treatment failure ($P=0.006$). This trend attenuated at day 180 (20 versus 16, $P=0.14$). Similarly, Treg:Tcon ratios at baseline and day 60 were higher among patients in whom treatment was a success than among those in whom it failed (*data not shown*).

B-cell numbers were not significantly different between arms, except at day 180 when the median B-cell number (cells per microliter) was higher in the two-drug arm, 148 (range, 0-1547) than in the three-drug arm, 68 (range, 0-1035) ($P=0.045$). Overall, median plasma BAFF levels were low (1.0 – 1.9 ng/mL) at all time-points, potentially

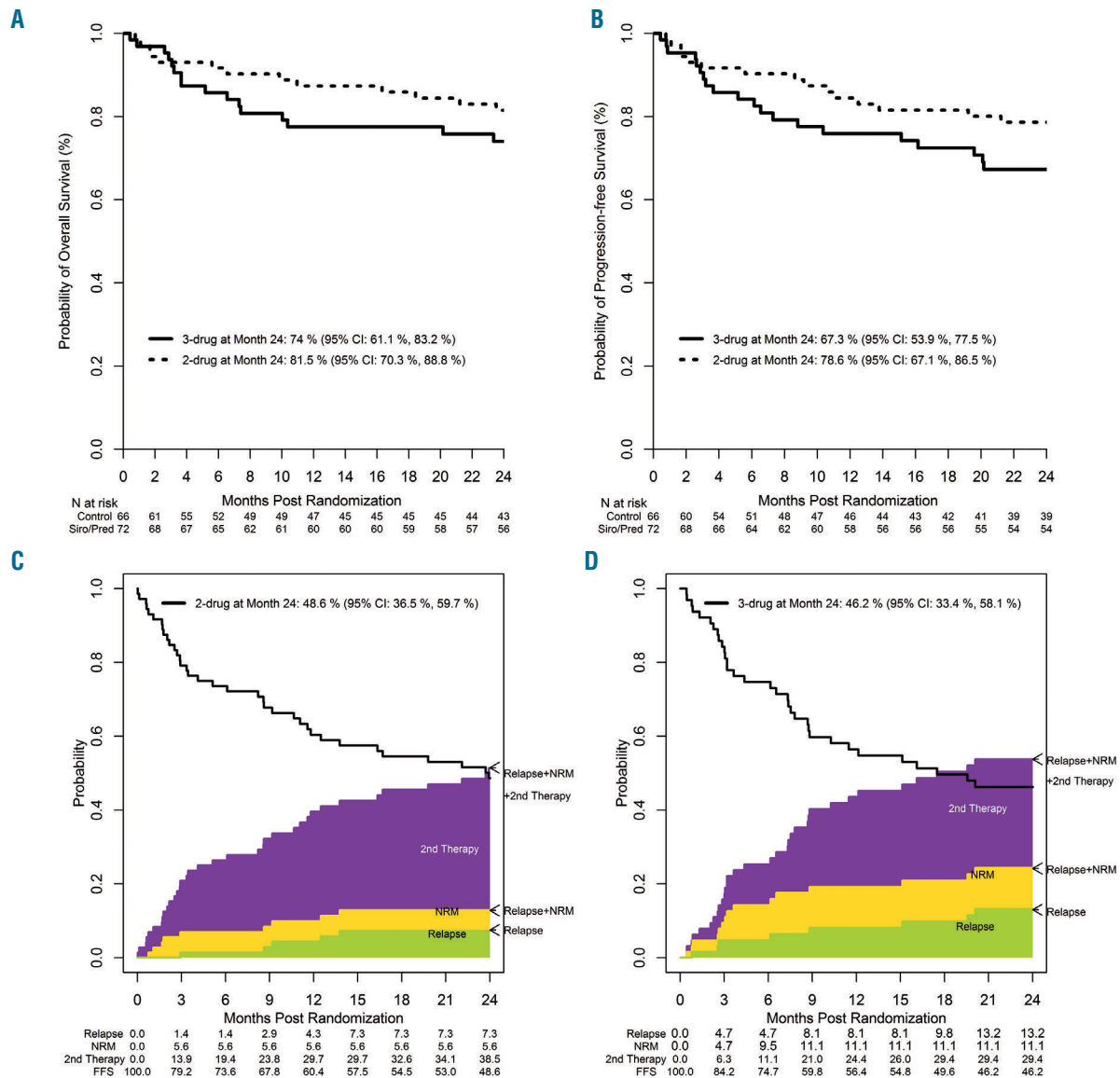


Figure 2. Overall survival, progression-free survival, and failure-free survival. (A) Probability of overall survival by treatment arm ($P=0.281$). (B) Probability of progression-free survival by treatment arm ($P=0.142$). Progression-free survival was defined as no clinical evidence of progression or relapsed disease, or any therapy used to treat persistent, progressive, or relapsed disease including withdrawal of immunosuppressive therapy or donor lymphocyte infusion. (C) Probability of failure-free survival in the (C) sirolimus + prednisone (2-drug) arm and (D) control (3-drug) arm. Failure-free survival was defined by the absence of secondary immunosuppressive therapy for chronic graft-versus-host disease, non-relapse mortality, and recurrent or progressive malignancy during treatment. Note: the numbers shown are estimates at each time point for each endpoint. Siro/Pred: sirolimus + prednisone; NRM: non-relapse mortality; FFS: failure-free survival.

Table 2. Chronic graft-versus-host disease characteristics.

	Two drugs ^a N = 72 (%)	Three drugs N = 66 (%)
Time from transplant to enrollment in months, median (range)	7.3 (3.4, 28.6)	7.7 (2.5, 42.4)
Time from diagnosis of chronic GvHD to enrollment in days, median (range)	7.0 (0.0, 202.0)	10.0 (0.0, 310.0)
Chronic GvHD risk status ^a		
High	37 (51.4)	32 (48.5)
Standard	35 (48.6)	34 (51.5)
Category of chronic GvHD diagnosis relative to enrollment		
Treatment naïve ^b	49 (68)	36 (54.5)
Early inadequate responders ^c	23 (32)	30 (46.5)
Chronic GvHD presentation		
Overlap acute and chronic GvHD	36 (50.0)	34 (53.1)
Classic chronic GvHD	36 (50.0)	30 (46.9)
Chronic GvHDs severity ^d		
Mild	16 (22.2)	13 (20.6)
Moderate	45 (62.5)	45 (71.4)
Severe	11 (15.3)	5 (7.9)
Chronic GvHD onset		
De-novo	30 (41.7)	24 (36.4)
Quiescent	10 (13.9)	14 (21.2)
Progressive	32 (44.4)	28 (42.4)
Skin score		
0	14 (19.7)	19 (29.7)
1	16 (22.5)	11 (17.2)
2	22 (31.0)	20 (31.3)
3	19 (26.8)	14 (21.9)
Mouth score		
0	20 (27.8)	16 (25.0)
1	31 (43.1)	26 (40.6)
2	18 (25.0)	17 (26.6)
3	3 (4.2)	5 (7.8)
Gastro-intestinal tract score		
0	40 (56.3)	33 (51.6)
1	17 (23.9)	21 (32.8)
2	12 (16.9)	8 (12.5)
3	2 (2.8)	2 (3.1)
Eye score		
0	33 (46.5)	29 (45.3)
1	25 (35.2)	24 (37.5)
2	9 (12.7)	10 (15.6)
3	4 (5.6)	1 (1.6)
Joints and fascia score		
0	53 (74.6)	49 (76.6)
1	10 (14.1)	11 (17.2)
2	7 (9.9)	4 (6.3)
3	1 (1.4)	0 (0.0)
Genital tract score		
0	43 (89.6)	28 (73.7)
1	3 (6.3)	8 (21.1)
2	2 (4.2)	0 (0.0)
3	0 (0.0)	2 (5.3)
Lung score		
0	62 (86.1)	54 (84.4)
1	8 (11.1)	6 (9.4)
2	1 (1.4)	3 (4.7)
3	1 (1.4)	1 (1.6)
Liver score		
0	19 (26.4)	21 (31.8)
1	14 (19.4)	23 (34.9)

*continued in the next column**continued from the previous column*

	Two drugs ^a N = 72 (%)	Three drugs N = 66 (%)
2	22 (30.6)	14 (21.2)
3	17 (23.6)	8 (12.1)
Acute GvHD prior to enrollment		
Yes	42 (58.3)	42 (63.6)
No	30 (41.7)	24 (36.4)

Two drugs: prednisone/sirolimus; three drugs: prednisone/sirolimus/calcineurin-inhibitor; GvHD: graft-versus-host disease. ^aChronic GvHD risk status was defined as high-risk in patients with platelets <100,000, >50% skin involvement, bronchiolitis obliterans or those receiving prednisone ≥0.5 mg/kg/day (or equivalent) at the time of chronic GvHD diagnosis. ^bTreatment-naïve: newly diagnosed for <14 days. ^cEarly inadequate responders: previously treated and not responding ≤16 weeks from time of first diagnosis of chronic GvHD. ^dChronic GvHD severity was defined as per the NIH Consensus criteria.³¹

because of steroid use,²³ and did not differ significantly at day 180 between patients in whom treatment had or had not been successful by day 180. Among patients in whom treatment was successful, the median baseline BAFF levels were slightly higher, 1.6 ng/mL (range, 0.2-17.7) *versus* 1.2 ng/mL (range, 0.1-19.4; $P=0.046$). BAFF/B-cell ratios were not significantly different between study arms or between patients in whom treatment was or was not successful.

Quality of life

Health-related quality of life was measured by self-report instruments, including the Functional Assessment of Cancer Therapy Bone Marrow Transplantation (FACT-BMT), Short Form-36 (SF-36), and Lee symptom scale. Figure 4 shows significantly better scores, at 2 and 6 months, with two drugs *versus* three drugs for the SF-36 Physical Component Summary, adjusted for baseline scores ($P=0.02$ and $P=0.04$), and demonstrated further by SF-36 Physical Functioning sub-scores ($P=0.007$ and $P=0.001$). Unadjusted, seven-item FACT-BMT Physical Well-Being scores were higher at 2 months with two drugs, but this observation did not hold when adjusted for baseline scores ($P=0.27$, *graph not shown*). Lee symptom scale scores did not differ between arms (*data not shown*).

Toxicity and adverse events

The proportion of patients with grade 3-5 toxicities was similar, 63.9% for those in the two-drug arm *versus* 56.1% in the three-drug arm. Thrombotic microangiopathy was observed in one patient (1.4%) in the two-drug arm and in three (4.5%) in the three-drug arm. The mean (standard deviation) increase in serum creatinine from baseline to 2 months was significantly higher with three drugs, 0.1 (0.3) mg/dL than with two drugs, 0 (0.2) mg/dL ($P=0.002$). The proportion of patients with serum creatinine >1.5 times baseline was significantly higher among patients treated with three drugs than among those treated with two drugs at both 2 months (11.7% *versus* 1.5%, $P=0.025$), and at 6 months (24.0% *versus* 7.8%, $P=0.016$). Twenty-seven subjects in the two-drug arm *versus* 27 in the three-drug arm (40.9% *versus* 37.5%, $P=0.682$) experienced severe to life-threatening/fatal infection episodes with slightly more infection episodes occurring in the three-drug arm (80 *versus* 52). One patient in the two-drug arm and another three in the three-drug arm developed non-infectious pneumonitis.

Discussion

In this phase II, multicenter, randomized trial comparing two CNI-free approaches against a CNI-containing comparator arm, the comparison of two *versus* three drugs showed similar outcomes with the CNI-free chronic GvHD therapy. Subjects who received prednisone/sirolimus had better renal and physical function at 2 and 6 months, but these improvements did not result in long-term advantages for recipients of the two-drug immunosuppressive therapy. However, between 6 and 12 months, 20%-30% of subjects treated with two drugs and 11%-24% treated with three drugs had already switched to secondary immunosuppressive therapy, potentially attenuating longer-term associations. Earlier studies suggested that narrower targeting of CNI and sirolimus blood levels might mitigate nephrotoxicity.^{9,25} Data to evaluate this were not collected, but it is conceivable that side effects, particularly with three drugs, were partially mitigated by detailed study guidance for managing CNI and sirolimus blood levels.

Table 3. Treatment success.

	Treatment success at 6 months		P-value
	Two drugs (N=72)	Three drugs (N=66)	
Treatment success ^a			0.871
Yes	35 (48.6%) [95% CI: 36.7%-60.7%]	33 (50.0%) [95% CI: 37.4%-62.6%]	
CR	3	6	
PR	32	27	
No	37 (51.4%) [95% CI: 39.3%-63.4%]	33 (50.0%) [95% CI: 37.4%-62.6%]	
Relapse	1	3	
Secondary therapy	14	7	
Death	4	6	
Not in CR/PR	18	17	

Treatment success at 2 years			
	Treatment success at 2 years		P-value
	Two drugs (N=72)	Three drugs (N=66)	
Treatment success ^b			0.899
Yes	10 (14.7%) [95% CI: 7.3%-25.4%]	9 (15.5%) [95% CI: 7.4%-27.4%]	
CR off IST	3	7	
CR on IST	7	2	
No	58 (85.3%) [95% CI: 74.6%-92.7%]	49 (84.5%) [95% CI: 72.6%-92.7%]	
VGPR ^c off IST	6	1	
VGPR ^c on IST	4	3	
PR off IST	1	3	
PR on IST	9	7	
Progression	1	2	
Relapse	5	8	
Secondary therapy	28	18	
Death ^d	4	7	
Not evaluable ^e	4	8	

Two drugs: prednisone/sirolimus, three drugs: prednisone/sirolimus/calcineurin-inhibitor; ^aTreatment success at 6 months is defined as a complete (CR) or partial response (PR) without secondary systemic immunosuppressive therapy (IST) and no recurrent malignancy or death through 6 months after randomization. ^bPartial response is excluded from the definition of treatment success at 2 years. ^cVery good partial response (VGPR) is defined as having trivial residual and asymptomatic graft-versus-host disease features, while on a prednisone dose that is physiological or less (defined as ≤ 5 mg daily or ≤ 10 mg every other day). This category splits further into truly off all IST (meaning also off sirolimus and/or calcineurin inhibitor or not). ^dDeath, relapse, and secondary therapy are mutually exclusive, based on which event occurred first. ^eParticipant withdrew study consent.

Our hypothesis was that CNI-free two-drug immunosuppressive therapy would not impede Treg expansion and would thereby abrogate chronic GvHD better than would three-drug immunosuppressive therapy. The few statistically significant associations that we observed in Treg number and BAFF serum levels were inconsistent and/or difficult to understand suggesting that chronic GvHD biomarker response associations are more complex than Treg alone. Rates of secondary therapy for lack of efficacy between day 90 and day 180 were almost double with two-drug immunosuppressive therapy compared to three-drug therapy, which might have attenuated predicted Treg response associations.

We learned important lessons to inform and potentially improve future study design in chronic GvHD. Had phase II been positive, the 51 additionally accrued subjects beyond the phase II target would have jump-started phase III accrual, thereby confirming “built-in” adaptive phase II/III design utility. Second, to begin endpoint review committee adjudications 3 years after enrollment and then discover that 10% of subjects had been ineligible was problematic. For a complex clinical syndrome like chronic GvHD, our fundamental eligibility criterion that required patients to be “diagnosed according to NIH guidelines” was too open to misinterpretation. This problem was resolved after introducing a screening checklist to confirm that patients met the NIH diagnostic criteria for chronic GvHD (*Online Supplementary Material*). Lastly, a similar problem existed with respect to complete/partial response evaluations. While considerable efforts have been made to standardize and develop more objective chronic GvHD response instruments,²⁶ case report forms for this trial had multiple sections with categorical check boxes and required measures in many organs. The completion of case report forms was inconsistent and has been considered burdensome by others.²⁷ Months of iterative communications between the endpoint review committee, site investigators, and the Data Coordinating Center to clean

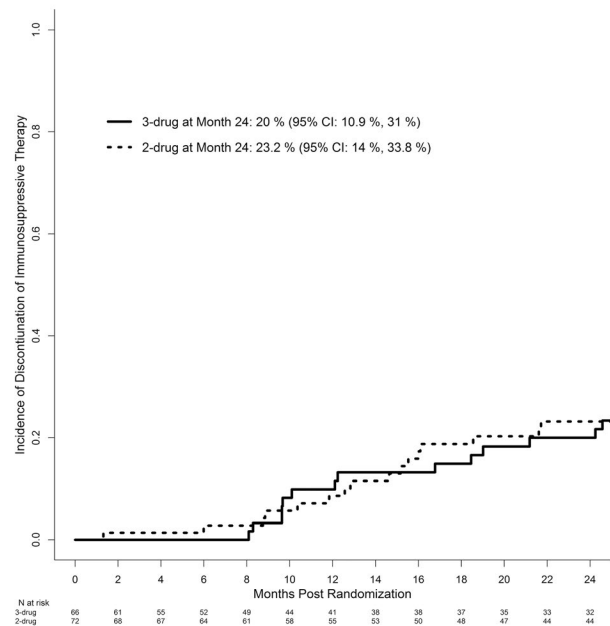


Figure 3. Cumulative incidence of discontinuation of all systemic immunosuppressive therapy without the need to add additional therapy by 2 years.

data were necessary to enable meaningful analysis. Near real-time monitoring (8-10 weeks after diagnosis) was invoked later by the BMT CTN in an effort to improve the quality of clinically annotated data in a more recent prospective correlative study of serum biomarkers, GvHD and other clinical outcomes.²⁸ These experiences from the 0801 CTN trial also directly informed simplifications implemented in the 2014 revision of the NIH diagnostic and response criteria.^{29,30}

Regardless of treatment arm, success rates at 6 months were only modest; approximately half of all subjects failed to achieve a complete or partial response and remain alive without relapse or receipt of secondary immunosuppressive therapy. There were fewer relapses in the two-drug arm, which contained more patients with early stage disease, more with acute leukemia (likely in first remission), and more recipients of myeloablative conditioning. Conceivably these chance imbalances might have resulted

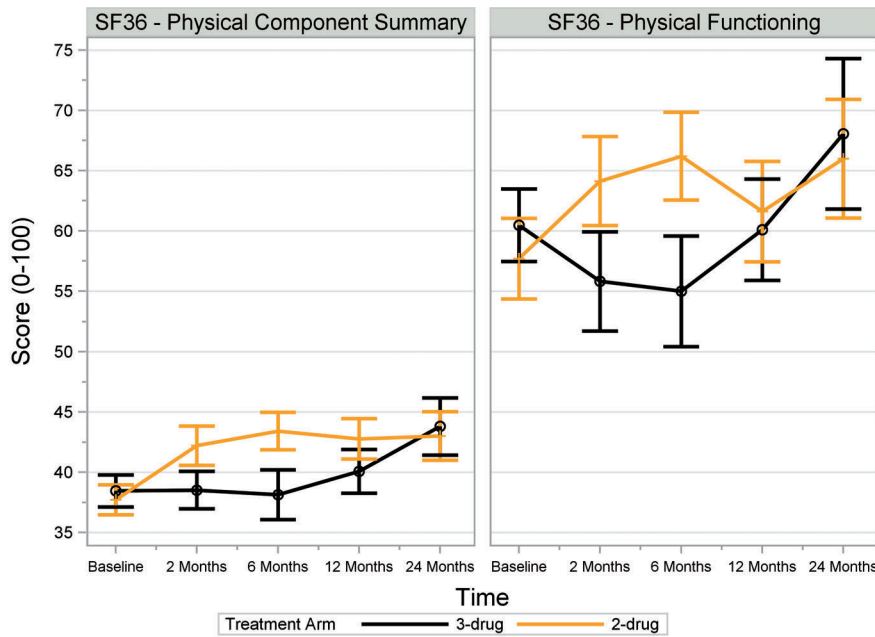


Figure 4. Quality of life. Unadjusted graphs are shown for SF-36 quality of life scores:²³ (left) Physical Component Summary scores; (right) Physical Functioning subscale scores.

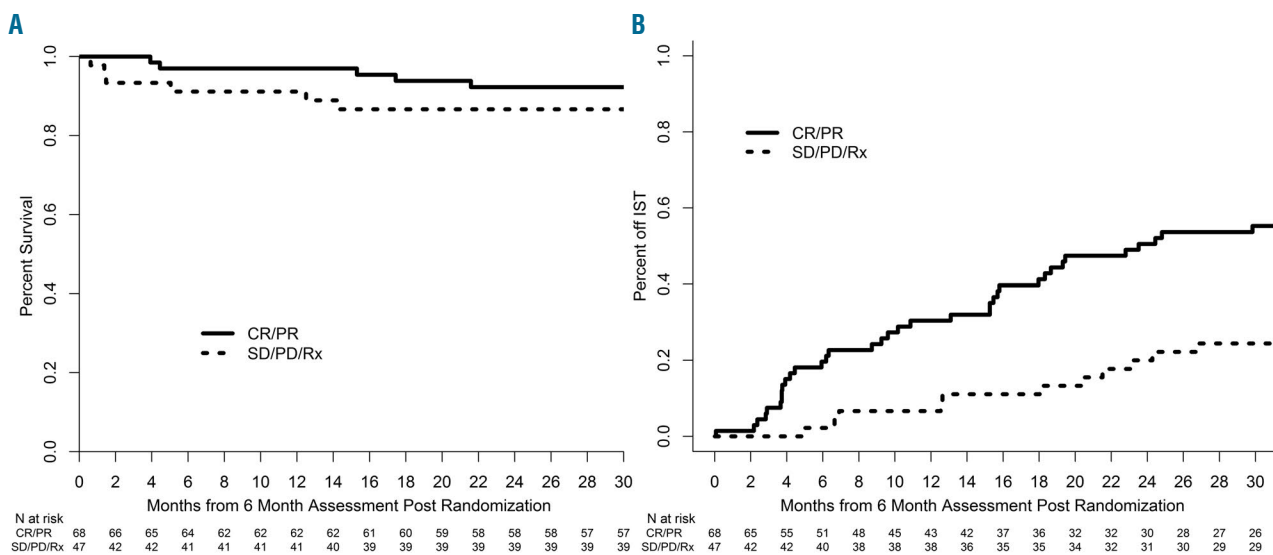


Figure 5. Six-month landmark analysis. (A) Overall survival after the 6-month landmark with a median follow up for survivors of 30 months. The mortality rates after the 6-month landmark were similar in the complete or partial response group and stable or progressive disease group (hazard ratio, 0.71; 95% confidence interval: 0.17-2.96; $P=0.63$) or secondary treatment groups (hazard ratio, 0.54; 95% confidence interval: 0.13-2.26; $P=0.39$), respectively. (B) Cumulative incidence of discontinuation of immunosuppressive therapy after the 6-month landmark. Patients who died or were lost to follow up or experienced relapse before the landmark were excluded ($n=19$) and no patients ended immunosuppressive therapy before the landmark. CR/PR: patients with complete or partial response without relapse and without secondary therapy at the time of assessment ($n=68$). SD/PD/Rx: patients not in complete or partial response and alive without relapse and without secondary therapy at the time of assessment ($n=31$), and patients who had secondary therapy without relapse regardless of response before the landmark ($n=16$); IST: immunosuppressive therapy.

in fewer relapses than in the potentially higher-risk patients who received three drugs. However, our sample size provided insufficient power to be able to detect a difference in relapse. The proportion of subjects who received secondary immunosuppressive therapy in either arm of the trial was not significantly different and it made no difference whether the patient had been enrolled with “newly diagnosed” versus “inadequately responding” chronic GvHD (*data not shown*).

Two large studies assessed failure-free survival among patients who received non-uniform initial chronic GvHD therapy and showed that the rates of failure-free survival were either similar³¹ or lower³² to those in our study in the context of overall generally higher rates of moderate and severe chronic GvHD. The Consortium recently evaluated 202 subjects in a landmark analyses to determine whether failure-free survival plus complete/partial responses at either 6 months or 1 year predicted downstream clinical benefit. At 1 year, the <20% of study subjects who satisfied the failure-free survival plus complete/partial response end-point versus all other states (failure-free survival with stable/progressive GvHD, or received secondary immunosuppressive therapy) were associated over the subsequent 5 years with significantly fewer GvHD disease manifestations, lower mortality, and earlier discontinuation of immunosuppressive therapy.³³ Because a similar 6-month failure-free survival plus complete/partial response landmark analysis revealed less striking associations with downstream clinical benefit, the authors proposed 1-year failure-free survival plus complete/partial responses as the primary endpoint for future pivotal clinical trials of initial therapy. Although 1-year endpoint data were not available, at 6 months our two groups without complete/partial response (i.e. failure-free survival with stable/progressive GvHD, or received secondary immunosuppressive therapy) behaved similarly when analyzed separately. They were, therefore, collapsed into one “stable disease/progressive disease/secondary immunosuppressive therapy” group of 47 patients for comparison to 68 patients with complete/partial response who satisfied the failure-free survival endpoint, resulting in just over half the size of the cohort used for the Consortium landmark analysis. Our 6-month failure-free survival plus complete/partial response endpoint predicted earlier time to discontinue immunosuppressive therapy compared to that in the combined group with stable or progressive disease or secondary immunosuppressive therapy (hazard ratio,

2.05; 95% CI: 1.15-3.68; $P=0.02$) (Figure 5), but similar to the findings of Martin *et al.*,³³ there was no striking survival benefit at 6 months, perhaps because of the shorter follow up and/or the small sample size. Larger prospective studies are needed to verify the utility of this endpoint in predicting better survival and shorter duration of immunosuppressive therapy.

In summary, this randomized trial showed no difference in response rates between the two treatment arms. Analyses of nephrotoxicity and quality of life demonstrate that initial therapy of chronic GvHD with prednisone/sirolimus is an acceptable alternative and better tolerated than three-drug therapy including a CNI. Our study could not address the relative merits of prednisone/sirolimus versus prednisone ± CNI. We were not able to include a prednisone ± CNI treatment arm (current standard for initial therapy), because the study included patients with high-risk chronic GvHD or early treatment failure. Success rates for prednisone/sirolimus as initial therapy in treatment-naïve or early inadequate responders are insufficient to warrant a randomized controlled trial versus prednisone with or without CNI. For early chronic GvHD therapy, novel approaches that improve rates of complete/partial responses and failure-free survival are required. Given the inherent complexity of chronic GvHD trials, we advise real-time diagnostic checklists to ensure patients' eligibility, and real-time data auditing to protect data integrity.

Acknowledgments

The authors would like to thank the National Heart, Lung, and Blood Institute and the National Cancer Institute for supporting this study (grant #U10HL069294). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank the laboratory of Jerome Ritz for performing all of the immunophenotyping and measuring serum BAFF levels. We thank Saurabh Chhabra, Gregory Yanik, Richard Maziarz, Suhag Parikh, Mark Litzow, Hillard Lazarus, Marcelo Pasquini, Andrew Artz, Krishna Gundabolu, Mark Juckett, Peter Westervelt, Pablo Parker, George Selby, George Chen, John Wingard, Scott Rowley, Scott Solomon, David Porter, Carlos Bachier, Paul Shaughnessy, James Essell, Marcie Riches, Thomas Shea, Michael Pulsipher, Edward Ball, John McCarty, Samantha Jaglowski, Guenther Koehne, and John Lister for enrolling patients on this trial. We thank the members of the Blood and Marrow Transplant Clinical Trials Network, the research nurses, and the patients who participated in this trial.

References

- Socie G, Stone JV, Wingard JR, et al. Long-term survival and late deaths after allogeneic bone marrow transplantation. Late Effects Working Committee of the International Bone Marrow Transplant Registry. *N Engl J Med*. 1999;341(1):14-21.
- Pidala J, Kurland B, Chai X, et al. Patient-reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. *Blood*. 2011;117(17):4651-4657.
- Sullivan KM, Witherspoon RP, Storb R, et al. Prednisone and azathioprine compared with prednisone and placebo for treatment of chronic graft-versus-host disease: prognostic influence of prolonged thrombocytopenia after allogeneic marrow transplantation. *Blood*. 1988;72(2):546-554.
- Koc S, Leisenring W, Flowers ME, et al. Therapy for chronic graft-versus-host disease: a randomized trial comparing cyclosporine plus prednisone versus prednisone alone. *Blood*. 2002;100(1):48-51.
- Arora M, Wagner JE, Davies SM, et al. Randomized clinical trial of thalidomide, cyclosporine, and prednisone versus cyclosporine and prednisone as initial therapy for chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2001;7(5):265-273.
- Koc S, Leisenring W, Flowers ME, et al. Thalidomide for treatment of patients with chronic graft-versus-host disease. *Blood*. 2000;96(12):3995-3996.
- Martin PJ, Storer BE, Rowley SD, et al. Evaluation of mycophenolate mofetil for initial treatment of chronic graft-versus-host disease. *Blood* 2009;113(21):5074-5082.
- Gilman AL, Schultz KR, Goldman FD, et al. Randomized trial of hydroxychloroquine for newly diagnosed chronic graft-versus-host disease in children: a Children's Oncology Group study. *Biol Blood Marrow Transplant*. 2012;18(1):84-91.

9. Couriel DR, Saliba R, Escalon MP, et al. Sirolimus in combination with tacrolimus and corticosteroids for the treatment of resistant chronic graft-versus-host disease. *Br J Haematol.* 2005;130(3):409-417.
10. Couriel DR, Hosing C, Saliba R, et al. Extracorporeal photochemotherapy for the treatment of steroid-resistant chronic GVHD. *Blood.* 2006;107(8):3074-3080.
11. Cutler C, Miklos D, Kim HT, et al. Rituximab for steroid-refractory chronic graft-versus-host disease. *Blood.* 2006;108(2):756-762.
12. Lopez F, Parker P, Nademanee A, et al. Efficacy of mycophenolate mofetil in the treatment of chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2005;11(4):307-313.
13. Jacobsohn DA, Chen AR, Zahurak M, et al. Phase II study of pentostatin in patients with corticosteroid-refractory chronic graft-versus-host disease. *J Clin Oncol.* 2007;25(27):4255-4261.
14. Zeiser R, Nguyen VH, Beilhack A, et al. Inhibition of CD4+CD25+ regulatory T-cell function by calcineurin-dependent interleukin-2 production. *Blood.* 2006;108(1):390-399.
15. Zeiser R, Levenson-Gower DB, Zambricki EA, et al. Differential impact of mammalian target of rapamycin inhibition on CD4+CD25+Foxp3+ regulatory T cells compared with conventional CD4+ T cells. *Blood.* 2008;111(1):453-462.
16. Coenen JJ, Koenen HJ, van Rijsen E, et al. Rapamycin, not cyclosporine, permits thymic generation and peripheral preservation of CD4+CD25+FoxP3+ T cells. *Bone Marrow Transplant.* 2007;39(9):537-545.
17. Gatza E, Rogers CE, Clouthier SG, et al. Extracorporeal photopheresis reverses experimental graft-versus-host disease through regulatory T cells. *Blood.* 2008;112(4):1515-1521.
18. Biagi E, Di Biaso I, Leoni V, et al. Extracorporeal photochemotherapy is accompanied by increasing levels of circulating CD4+CD25+GITR+Foxp3+CD62L+ functional regulatory T-cells in patients with graft-versus-host disease. *Transplantation.* 2007;84(1):31-39.
19. Filipovich AH, Weisdorf D, Pavletic S, 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(12):945-956.
20. Pavletic SZ, Martin PJ, Lee SJ, et al. 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(3):252-266.
21. Martin PJ, Weisdorf D, Przepiorka D, 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(5):491-505.
22. Carpenter PA, Kitko CL, Elad S, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: V. The 2014 ancillary therapy and supportive care working group report. *Biol Blood Marrow Transplant.* 2015;21(7):1167-1187.
23. Sarantopoulos S, Stevenson KE, Kim HT, et al. Altered B-cell homeostasis and excess BAFF in human chronic graft-versus-host disease. *Blood.* 2009;113(6):3865-3874.
24. Ware JE, Kosinski M, Bjorner JB, et al. User's Manual for the SF-36v2 Health Survey. 2nd edn. QualityMetric Incorporated; Lincoln, RI: 2007.
25. Johnston LJ, Brown J, Shizuru JA, et al. Rapamycin (sirolimus) for treatment of chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2005;11(1):47-55.
26. Lee SJ. Classification systems for chronic graft-versus-host disease. *Blood.* 2017;129(1):30-37.
27. Duarte RF, Greinix H, Rabin B, et al. Uptake and use of recommendations for the diagnosis, severity scoring and management of chronic GVHD: an international survey of the EBMT-NCI Chronic GVHD Task Force. *Bone Marrow Transplant.* 2014;49(1):49-54.
28. <https://clinicaltrials.gov/ct2/show/NCT01879072>
29. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. The 2014 diagnosis and staging working group report. *Biol Blood Marrow Transplant.* 2015;21(3):389-401.
30. Lee SJ, Wolff D, Kitko C, 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. The 2014 response criteria working group report. *Biol Blood Marrow Transplant.* 2015;21(6):984-99.
31. Inamoto Y, Flowers ME, Sandmaier BM, et al. Failure-free survival after initial systemic treatment of chronic graft-versus-host disease. *Blood.* 2014;124(8):1363-1371.
32. Palmer J, Chai X, Martin PJ, et al. Failure-free survival in a prospective cohort of patients with chronic graft-versus-host disease. *Haematologica.* 2015;100(5):690-695.
33. Martin PJ, Storer BE, Inamoto Y, et al. An endpoint associated with clinical benefit after initial treatment of chronic graft-versus-host disease. *Blood.* 2017;130(3):360-367.