

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

Joseph Pidala,¹ Brenda Kurland,² Xiaoyu Chai,² Navneet Majhail,³ Daniel J. Weisdorf,³ Steven Pavletic,⁴ Corey Cutler,⁵ David Jacobsohn,⁶ Jeanne Palmer,⁷ Sally Arai,⁸ Madan Jagasia,⁹ and Stephanie J. Lee²

¹Moffitt Cancer Center, Tampa, FL; ²Fred Hutchinson Cancer Research Center, Seattle, WA; ³University of Minnesota, Minneapolis, MN; ⁴National Cancer Institute, Bethesda, MD; ⁵Dana-Farber Cancer Institute, Boston, MA; ⁶Children's Memorial Hospital, Chicago, IL; ⁷Medical College of Wisconsin, Milwaukee, WI; ⁸Stanford University, Stanford, CA; and ⁹Vanderbilt University, Nashville, TN

Quality of life (QOL) after hematopoietic cell transplantation (HCT) is compromised by chronic GVHD. In a prospectively assembled multicenter cohort of adults with chronic GVHD (n = 298), we examined the relationship between chronic GVHD severity defined by National Institutes of Health (NIH) criteria and QOL as measured by the SF-36 and FACT-BMT instruments at time of enrollment. Chronic GVHD severity was independently associated with QOL, adjusting for age. Compared with population

normative data, SF-36 scores were more than a SD (10 points) lower on average for the summary physical component score (PCS) and role-physical subscale, and significantly lower (with magnitude 4-10 points) for several other subscales. Patients with moderate and severe cGVHD had PCS scores comparable with scores reported for systemic sclerosis, systemic lupus erythematosus, and multiple sclerosis, and greater impairment compared with common chronic conditions including diabetes, hypertension,

and chronic lung disease. Moderate to severe cGVHD as defined by NIH criteria is associated with significant compromise in multiple QOL domains, with PCS scores in the range of other systemic autoimmune diseases. Compromised QOL provides a functional assessment of the effects of chronic GVHD, and may be measured in cGVHD clinical studies using either the SF-36 or the FACT-BMT. (*Blood*. 2011;117(17):4651-4657)

Introduction

Quality of life (QOL) is a multi-dimensional construct composed of several related domains including physical, emotional, social, and role functioning, as well as a person's overall evaluation of his or her well being and ability to function.¹⁻⁴ Previous studies have demonstrated the adverse impact of chronic GVHD on QOL.⁵⁻¹⁰ As QOL is one of transplant survivors' central concerns, its study is of vital importance.

Chronic GVHD represents the most important source of late nonrelapse mortality after HCT.^{11,12} This syndrome is responsible for significant morbidity, impaired functional status, and prolonged duration of immunosuppression after HCT.¹²⁻¹⁴ In a series of publications originating from the 2004 National Institutes of Health (NIH) Consensus Conference, investigators proposed means to standardize diagnosis, scoring, histopathology, biomarkers, response assessment, and the conduct of clinical trials in chronic GVHD.¹⁵⁻²⁰ These criteria were developed to advance clinical trials in chronic GVHD. As QOL is an essential measure in the patients' and physicians' evaluation of treatment outcome, it should be subjected to the same degree of rigorous study as other relevant treatment outcomes. An understanding of the relationship between QOL and chronic GVHD severity and response to treatment is necessary to facilitate conduct of clinical trials for chronic GVHD prevention and treatment.

Accordingly, we have examined QOL according to chronic GVHD severity defined by NIH consensus scoring in baseline data

for the Chronic GVHD Consortium, a prospectively assembled cohort of chronic GVHD affected HCT recipients. The aims of this study are to (1) describe the relationship between chronic GVHD severity and patient-reported QOL; (2) compare QOL in HCT recipients with chronic GVHD to US population normative data; (3) compare QOL in HCT recipients with chronic GVHD to patients with other chronic health conditions; and (4) investigate the ability of SF-36 and FACT-BMT QOL instruments to discriminate chronic GVHD severity.

Methods

Chronic GVHD Consortium: description of study cohort and cohort for this analysis

A cohort of HCT recipients with chronic GVHD was prospectively assembled in a multicenter observational study. The protocol was approved by the Institutional Review Board at each of the 5 sites (Fred Hutchinson Cancer Research Center, University of Minnesota, Dana-Farber Cancer Institute, Stanford University, and Vanderbilt University), and all subjects provided informed consent in accordance with the Declaration of Helsinki. Patients enrolled in the cohort were allogeneic HCT recipients age 2 or greater with chronic GVHD requiring systemic immunosuppressive therapy, including both those with classic chronic GVHD and those with overlap syndrome.¹⁹ Cases were classified as incident (enrollment < 3 months after

Submitted November 15, 2010; accepted February 6, 2011. Prepublished online as *Blood* First Edition paper, February 25, 2011; DOI 10.1182/blood-2010-11-319509.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2011 by The American Society of Hematology

chronic GVHD diagnosis) or prevalent (enrollment 3 or more months but < 3 years after chronic GVHD diagnosis). Primary disease relapse, and inability to comply with study procedures were exclusion criteria. At enrollment and every 6 months thereafter, physicians and patients report standardized information on chronic GVHD organ involvement and symptoms. Incident cases had an extra assessment time point 3 months after enrollment. Chronic GVHD severity was calculated from individual organ scoring provided by clinicians using the NIH consensus scoring (mild, moderate, severe).¹⁹ Standardized chart review after each visit abstracted objective medical data (including ancillary testing and laboratory results), medical complications, and medication profiles. This analysis only examines adult (age 18+ years) patients' QOL. Only the baseline (time of enrollment) data are analyzed for patients enrolled as of December 2009; enrollment and collection of longitudinal data are ongoing.

QOL instruments

The FACT-BMT and the SF-36 were administered to assess patient-reported QOL. The FACT-BMT Version 4.0 is a 37 item self-report questionnaire, which includes a 10 item Bone Marrow Transplant Subscale (BMTS). The instrument measures the effect of cancer therapy on multiple QOL domains including physical (PWB), functional (FWB), social/family, and emotional well being, and BMT-specific concerns. Individual domain scores are summarized to give a total FACT-BMT score. As well, a FACT-TOI (trial outcome index) score consists of the sum of physical and functional well being and the BMT subscale (PWB + FWB + BMTS).^{21,22}

The SF-36 Version 2 is a 36-item self-report questionnaire which assesses patient-reported health and functioning. The instrument examines the following domains of QOL: physical functioning (PF), role functioning-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role functioning-emotional (RE), and mental health (MH). Two summary scales from the SF-36 include the physical component score (PCS) and the mental component score (MCS).²³⁻²⁷

Statistical methods

Standard algorithms were used to compute total and subscale scores for FACT-BMT²¹ and SF-36 instruments.^{24,25} Graphical displays and linear correlation were used to describe the relationships between individual QOL domains within the FACT-BMT and the SF-36. Subjects' QOL scores were displayed by chronic GVHD severity (mild, moderate, severe) according to NIH consensus criteria for global severity. In univariate analysis, the relationship between chronic GVHD severity and patient-reported QOL was examined using NIH chronic GVHD severity as the independent variable of interest. Linear-mixed models were planned to account for a random effect of study site. However, the variance of this random intercept was estimated as near the boundary value (0) for all models indicating no effect of transplant center, so the results shown are for linear regression models without random effects. Linear contrasts were used to estimate pairwise differences in average QOL scores between mild/moderate, moderate/severe, and mild/severe chronic GVHD severity.

Multivariable models were constructed to determine whether GVHD severity was associated with QOL after controlling for patient and disease characteristics. Covariates considered were age at enrollment, time from HCT to enrollment, disease stage (early/intermediate/advanced), donor type (matched sibling donor vs other), conditioning regimen (myeloablative vs not), chronic GVHD status at enrollment (incident vs prevalent case), subject gender, and education level (5-level scale). Separate models were developed for each QOL composite or subscale score for both the SF-36 and FACT-BMT instruments, although each covariate carried forward to the multivariable model with severity was included in multivariable models for all subscales. Statistical interaction (effect modification) was not investigated, because no interactions were expected and any findings would likely be spurious with small subgroups created by the interaction terms.

To quantify the magnitude of impairment in QOL in chronic GVHD affected subjects, we compared SF-36 total and subscale mean scores (according to chronic GVHD severity per the NIH score) to age- and

sex-adjusted US population normative mean SF-36 scores. First, individual scores were subtracted from age- and sex-specific means. These differences were evaluated using a sign test of the null hypothesis that each patient's score is equally likely to be higher or lower than the age and gender adjusted norm. Means and 95% confidence intervals for SF-36 scores of chronic GVHD cohort subjects were compared graphically to means and 95% confidence intervals for SF-36 scores reported for selected chronic health conditions.

One aim of the observational study is to evaluate measures of QOL in chronic GVHD patients to determine which would be most useful in evaluating status and change in GVHD symptom burden in clinical trials. We therefore compared the association between QOL and severity measures through an extension of receiver operating characteristic (ROC) methods and the concordance index for an ordinal "gold standard" (chronic GVHD severity) classified by levels of a marker (QOL measures).^{28,29}

Statistical analyses were conducted using SAS/STAT software, Version 9.2 (SAS Institute) and R Version 2.9.2 (R Foundation for Statistical Computing). To recognize the multiple tests introduced by comparing several QOL measures and pairwise comparisons among GVHD severity levels, type I error was controlled by considering a *P* value of 0.01 or lower as statistically significant, and by looking for consistency of results across related constructs.

Results

Chronic GVHD characteristics and baseline QOL scores

A total of 298 subjects meeting analysis criteria were enrolled between August 2007 and December 2009 at 5 centers. Enrollment at each site included Fred Hutchinson Cancer Research Center (n = 158, or 53%), Stanford (n = 48, or 16%), University of Minnesota (n = 35, or 12%), Dana-Farber Cancer Institute (n = 34, or 11%), and Vanderbilt (n = 23, or 8%). NIH severity was mild in 31 (10%), moderate in 175 (59%), and severe in 92 (31%). Median age was 53 (range 20-79). Overall the cohort was 92% white, 58% male with 57% receiving myeloablative conditioning, and 89% peripheral blood. HLA-identical siblings were the donor in 46% of cases and unrelated donors in 51% of cases. Approximately half (54%) of the cases were diagnosed within the previous 3 months. Both acute and chronic GVHD manifestations (overlap syndrome) were present in 44% of cases while 56% had only classic chronic manifestations.

Two hundred sixty subjects (87%) completed all or part of the SF-36 and FACT-BMT questionnaires. The other 38 patients (13%) were missing their entire patient questionnaires at baseline. Of those who completed the SF-36, individual items were completed 97%-100% of the time except for sexual functioning, which allowed patients opt out of answering the question if they were not sexually active and a question about concern about keeping a job, which would not be relevant to someone retired or unemployed (see supplemental Data; available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). Reasons for missing surveys most commonly included patient not returning a survey despite 3 attempts to collect it (55%), patient too ill to complete (16%), and patient refusal (8%). The rate of missing questionnaires was similar for patients with mild (5/31, 16%), moderate (22/175, 13%), and severe GVHD (11/92, 12%). Enrollment SF-36 and FACT-BMT total and subscale scores are described according to NIH chronic GVHD severity in Table 1.

To characterize the relationship between QOL domains, Pearson correlation coefficients were calculated for all pairings of SF-36 and FACT-BMT subscales. In the SF-36, theoretically related constructs demonstrated high correlation (physical/role

Table 1. QOL scores for cohort members at enrollment according to NIH chronic GVHD severity

	All (N = 298)			Mild (n = 31)			Moderate (n = 175)			Severe (n = 92)		
	n (%)	n	Median (range)	n (%)	n	Median (range)	n (%)	n	Median (range)	n (%)	n	Median (range)
SF-36*												
Physical-functioning scale		256	42.3 (14.9-57.0)		26	46.5 (23.4-57.0)		151	42.3 (17.0-57.0)		79	38.1 (14.9-57.0)
Role-functioning physical scale		259	37.3 (17.7-56.9)		26	44.6 (17.7-56.9)		153	37.3 (17.7-56.9)		80	32.4 (17.7-56.9)
Bodily pain scale		259	46.1 (19.9-62.1)		26	46.1 (24.9-62.1)		152	46.1 (24.1-62.1)		81	41.4 (19.9-62.1)
General health perceptions scale		259	41.0 (18.6-63.9)		26	46.5 (28.1-63.9)		152	41.0 (18.6-62.5)		81	37.7 (23.4-62.5)
Vitality scale		260	45.8 (20.9-70.8)		26	50.5 (30.2-64.6)		153	45.8 (20.9-70.8)		81	45.8 (20.9-70.8)
Social-functioning scale		260	40.5 (13.2-56.8)		26	45.9 (13.2-56.8)		153	40.5 (13.2-56.8)		81	35.0 (13.2-56.8)
Role-functioning emotional scale		259	48.1 (9.2-55.9)		26	52.0 (28.7-55.9)		153	48.1 (9.2-55.9)		80	44.2 (9.2-55.9)
Mental health scale		260	52.8 (13.4-64.1)		26	54.2 (30.3-64.1)		153	52.8 (13.4-64.1)		81	50.0 (21.8-64.1)
Physical component score		254	39.2 (15.9-59.4)		26	43.2 (19.9-59.3)		149	40.2 (15.9-59.4)		79	36.3 (19.3-56.7)
Mental component score		254	51.0 (15.3-68.4)		26	53.8 (28.5-61.4)		149	51.4 (17.0-68.4)		79	48.1 (15.3-65.8)
FACT-BMT												
Physical well-being		259	22.0 (4.0-28.0)		26	24.0 (9.0-28.0)		152	22.5 (4.0-28.0)		81	19.0 (7.0-28.0)
Social/family well-being		259	23.3 (0-28.0)		26	24.0 (11.0-28.0)		152	24.0 (0-28.0)		81	23.0 (6.0-28.0)
Emotional well-being		258	20.0 (6.0-24.0)		26	20.5 (14.0-24.0)		151	20.0 (6.0-24.0)		81	19.0 (6.0-24.0)
Functional well-being		258	17.0 (3.0-28.0)		26	17.0 (7.0-28.0)		152	18.0 (5.0-28.0)		80	15.0 (3.0-28.0)
BMT subscale		258	27.6 (11.0-39.0)		26	29.0 (14.0-36.0)		152	28.0 (11.0-39.0)		80	26.0 (11.1-38.9)
FACT total		257	81.0 (29.0-107.0)		26	84.5 (55.0-107.0)		151	84.1 (29.0-104.0)		80	73.4 (37.0-105.0)
FACT trial outcome index		258	65.0 (22.0-92.0)		26	70.0 (38.0-91.0)		152	67.8 (22.0-92.0)		80	58.5 (25.0-89.9)
FACT-BMT total		257	109.0 (48.1-142.0)		26	114.0 (69.0-142.0)		151	111.0 (49.0-140.0)		80	98.7 (48.1-136.0)

QOL indicates quality of life.

*Norm-based scores (mean = 50, SD = 10 based on 1998 general US population).

functioning-physical, $r = 0.70$; PCS/physical, $r = 0.84$; PCS/role functioning-physical, $r = 0.81$; MCS/mental health, $r = 0.87$). Conversely, there was minimal correlation between the MCS and PCS ($r = 0.23$). For the FACT-BMT, small correlation was observed in theoretically dissimilar domains (FACT-TOI/social-family, $r = 0.46$; PWB/social-family, $r = 0.32$; PWB/emotional, $r = 0.58$; FWB/emotional, $r = 0.56$; FWB/social-family, $r = 0.44$).

Impact of chronic GVHD severity on patient-reported QOL

Of the covariates considered for multivariable analysis, only age demonstrated an association with QOL scales, and only with the SF-36 physical functioning. Age was included as a covariate in all multivariable models, with age and GVHD severity predicting levels of QOL composite and subscale scores (Table 2). Statistically significant differences were found most often between

average QOL scores in moderate and severe GVHD severity categories, with fewer significant differences between mild and moderate GVHD severity. However, these comparisons reflect the small sample size for mild GVHD by NIH criteria ($n = 31$), as well as smaller magnitude of effects (all differences in average QOL were < 5 points for mild vs moderate).

QOL composite and subscale scores were higher on average for patients with moderate GVHD compared with severe GVHD. The estimated average difference for the SF-36 role-physical was 4.17 points (95% confidence interval 1.13, 7.22), compared with a SD of 10 points for an unselected population. Statistically significant differences were also observed for the FACT total, TOI, and BMT subscale. As expected, differences in QOL between mild and severe GVHD groups were of greater magnitude than between moderate and severe. Figure 1 shows these results graphically, with

Table 2. Multivariable models examining QOL outcomes

	Age				Mild vs moderate				Moderate vs severe				Mild vs severe			
	Estimate	95% lower	95% upper	P	Estimate	95% lower	95% upper	P	Estimate	95% lower	95% upper	P	Estimate	95% lower	95% upper	P
PCS	-0.11*	-0.21*	-0.004*	.04*	2.41	-1.57	6.39	.23	3.20*	0.60*	5.81*	.02*	5.61†	1.38†	9.84†	.009†
MCS	0.07	-0.04	0.18	.20	3.28	-1.11	7.67	.14	2.16	-0.72	5.04	.14	5.44*	0.77*	10.11*	.02*
FACT trial outcome index	0.005	-0.15	0.16	.95	3.19	-3.06	9.43	.32	5.53*	1.47*	9.60*	.01*	8.72*	2.08*	15.36*	.01*
Physical functioning	-0.20†	-0.31†	-0.09†	.0006†	2.10	-2.29	6.50	.35	3.06*	0.18*	5.93*	.04*	5.16*	0.48*	9.84*	.03*
Role-physical	-0.06	-0.18	0.06	.29	4.31	-0.38	8.99	.07	4.17†	1.13†	7.22†	.008†	8.48†	3.50†	13.46†	.0009†
Bodily pain	0.06	-0.05	0.18	.28	1.59	-2.91	6.09	.49	2.81	-0.11	5.72	.06	4.40	-0.38	9.17	.07
General health	0.006	-0.09	0.11	.91	3.11	-0.86	7.07	.12	2.86*	0.29*	5.43*	.03*	5.97†	1.76†	10.17†	.006†
Vitality	-0.10	-0.21	0.01	.09	2.79	-1.68	7.25	.22	3.36*	0.47*	6.26*	.02*	6.15*	1.40*	10.90*	.01*
Social functioning	0.02	-0.11	0.15	.78	2.08	-3.01	7.17	.42	4.08*	0.79*	7.38*	.02*	6.17*	0.76*	11.58*	.03*
Role-emotional	0.005	-0.13	0.14	.95	4.35	-0.95	9.65	.11	3.35	-0.09	6.80	.06	7.70*	2.07*	13.34*	.007*
Mental health	0.07	-0.04	0.17	.20	2.92	-1.19	7.03	.16	1.65	-1.01	4.32	.22	4.57*	0.20*	8.94*	.04*
FACT total	0.07	-0.09	0.24	.37	2.74	-3.68	9.15	.40	5.94†	1.76†	10.12†	.006†	8.68*	1.86*	15.50*	.01*
FACT-BMT total	0.06	-0.15	0.27	.59	4.37	-4.04	12.77	.31	7.77†	2.30†	13.24†	.006†	12.14†	3.20†	21.07†	.008†

Multivariable models examining QOL outcomes (SF-36 summary PCS and MCS, and subscales; FACT trial outcome index, FACT total, and FACT-BMT total) according to chronic GVHD severity.

MCS indicates mental component score; QOL, quality of life; and PCS, physical component score.

* $P < .05$.

† $P < .01$.

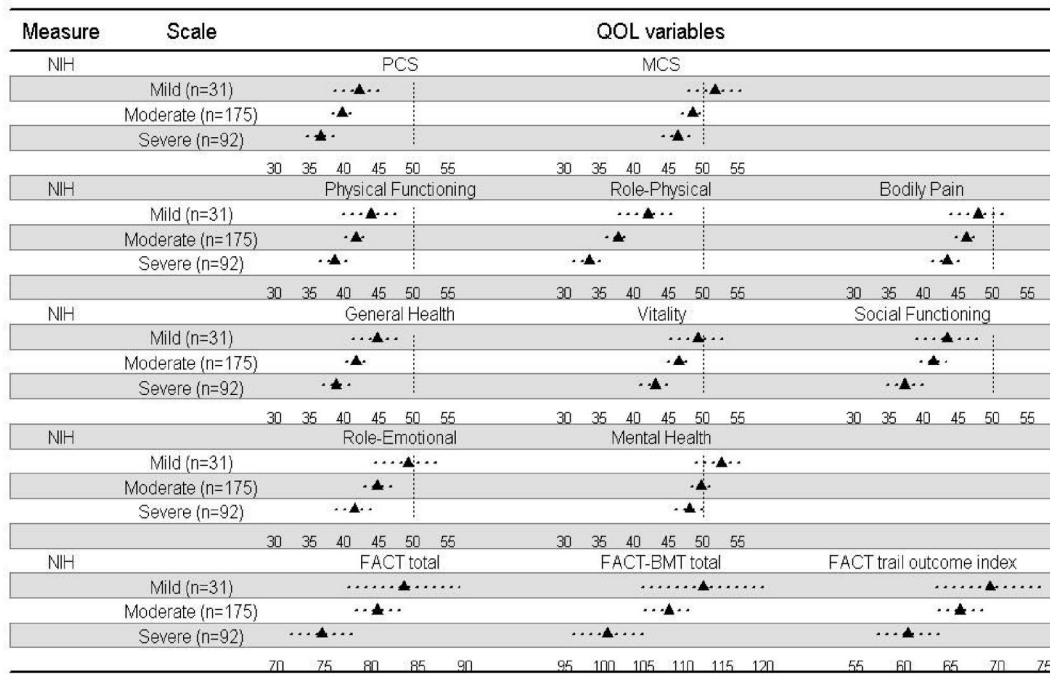


Figure 1. Fitted average QOL values and 95% confidence intervals for a 51-year-old with mild, moderate, or severe GVHD according to NIH criteria severity. Normal population mean is 50 (vertical dotted line) for SF-36 subscales.

fitted average QOL and 95% confidence interval for the mean for chronic GVHD severity levels according to the NIH criteria severity score. Age is held constant at the average value of about 51 years.

Comparison to population normative data

Figure 1 also shows the population norm for SF-36 subscales (50 points), marked as a vertical line, for comparison to fitted average QOL scores. To quantify the magnitude of impairment in QOL, we compared chronic GVHD cohort members' SF-36 mean scores to age- and gender-matched US population normative data. Mean scores for chronic GVHD cohort members were significantly lower for physical functioning, role-physical, bodily pain, general health, vitality, social functioning, and PCS. There were no significant differences observed in the domains of role-emotional, mental health, or MCS (Table 3).

Comparison to chronic health conditions

To further ascertain the clinical magnitude of QOL impairment observed in chronic GVHD cohort members, mean SF36 scores (PCS and MCS) of chronic GVHD cohort members according to NIH severity criteria were compared with those of other chronic health conditions.^{24,25,30-34} As demonstrated in Figure 2, those with moderate to severe chronic GVHD (rows 4 and 5) had decrement from expected population normative PCS scores comparable with that previously reported for systemic sclerosis, systemic lupus erythematosus, and multiple sclerosis (rows 6-9), but greater impairment compared with several common chronic health conditions including chronic lung disease, hypertension, diabetes and arthritis. Patients with mild or moderate chronic GVHD had MCS scores in keeping with population normative data and similar to the

Table 3. Comparison of mean SF-36 scores between chronic GVHD (cGVHD) cohort members and US population normative data

	cGVHD patients* (n = 260)		Normal population† (n = 260)		Difference score (cGVHD patient minus age-sex-matched expected normal value)		P‡
	Mean	SD	Mean	SD	Mean	SD	
Physical functioning	40.94	10.83	49.56	2.84	-8.61	10.65	< .0001
Role-physical	36.80	11.46	49.71	2.24	-12.91	11.61	< .0001
Bodily pain	45.44	10.83	49.50	1.48	-4.05	11.03	.0042
General health	41.10	9.59	49.71	1.11	-8.60	9.68	< .0001
Vitality	45.65	10.86	50.77	1.50	-5.11	11.12	< .0001
Social functioning	40.28	12.30	50.15	0.85	-9.87	12.32	< .0001
Role-emotional	44.20	12.80	50.25	1.14	-6.04	12.84	.0621
Mental health	49.47	9.90	50.67	1.77	-1.21	9.97	.1536
PCS	38.99	9.69	49.33	2.67	-10.34	9.84	< .0001
MCS	48.05	10.59	50.83	1.83	-2.78	10.65	.6606

*Norm-based scores (mean = 50, SD = 10 based on the 1998 general US population).

†Expected values for each cGVHD patient based on age and sex norm averages for the 1998 general US population. Expected physical functioning is slightly lower than 50 on average, and more variable for the patients, presumably due to associations with age.

‡Sign test (null hypothesis that it is equally likely that each cGVHD patient's score will be higher or lower than the age and sex norm).

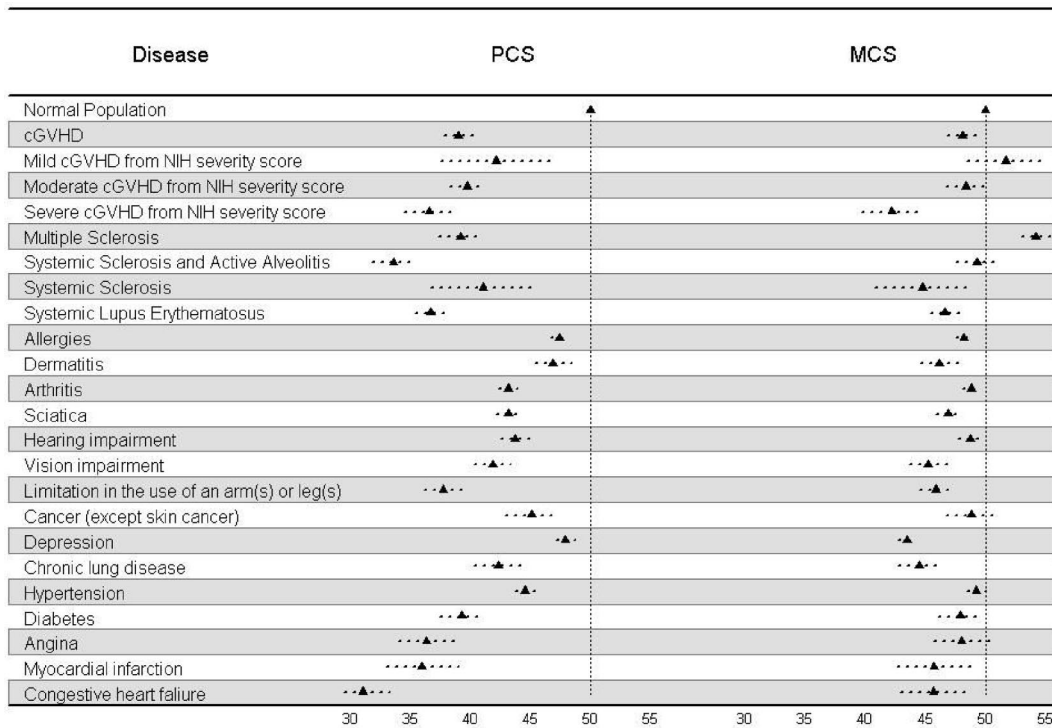


Figure 2. Comparison of SF-36 PCS and MCS mean scores (and 95% confidence intervals for the mean) from chronic GVHD cohort members according to NIH severity score and chronic health conditions. Normal population mean is 50 (vertical dotted line).

reported chronic health conditions. Interestingly, those with severe chronic GVHD had MCS scores comparable with depression.

Comparison of FACT-BMT and SF36 instruments

The graphical displays in Figure 2 demonstrate that, while average QOL differs by GVHD severity, there is overlap of QOL values between levels of GVHD severity for all QOL measures. This graphical result is supported by the findings of the diagnostic accuracy analysis (Table 4). Taking all possible pairings of patients and excluding ties (same GVHD severity), the area under the ROC curve (AUC) is the proportion of pairs for which the QOL measure for the patient with less severe GVHD is higher than the QOL measure for the patient with more severe GVHD. The concordance index was modest (~ 0.60) for all QOL scales examined. Using estimated variance and covariance of the AUCs to test the null hypothesis of no difference in accuracy,²⁸ we conclude that there were no significant differences between QOL instruments’ ability to discriminate between levels of chronic GVHD severity. Weighting schemes to allow less penalty for a 1-level difference (QOL measure for patient with moderate GVHD higher than for patient

with mild GVHD, or severe higher than moderate) than for a 2-level difference (severe higher than mild) did not affect the conclusion that the QOL measures compared did not differ in performance for classifying GVHD severity.

Discussion

QOL is routinely cited by cancer patients as a concern of central importance. Chronic GVHD threatens QOL after HCT, with previous studies demonstrating moderate to large impairments in multiple domains of QOL compared with those not affected by chronic GVHD.⁵⁻¹⁰ However, the impact of chronic GVHD severity according to the proposed NIH consensus criteria on patient-reported QOL among a cohort of exclusively chronic GVHD affected HCT recipients has not been examined to date. We report the baseline QOL data of chronic GVHD affected HCT recipients at the time of enrollment in the Chronic GVHD Consortium, a multicenter, prospective observational cohort study.

Several important findings emerge from this analysis. First, we have demonstrated that chronic GVHD severity according to the NIH criteria is significantly associated with patient-reported QOL, independent of other disease, transplantation, and socio-demographic variables. This effect was observed across multiple domains of QOL, indicating a wide-reaching impact on chronic GVHD patients’ reported QOL. Interestingly, none of the examined covariates, excepting the impact of age on physical functioning, was significantly associated with patient-reported QOL. Of particular relevance, we did not detect differences according to chronic GVHD status (incident vs prevalent) or time from HCT to enrollment: controlling for age at enrollment, the average PCS was estimated as only 0.3 points higher for incident than for prevalent cases, and not statistically different from no difference (*P* = .82). This is of importance, as the anticipated normal trajectory after

Table 4. Diagnostic accuracy analyses for GVHD severity, comparing AUC for the SF-36 PCS to the MCS and to FACT summaries

	N	AUC (SE)	<i>P</i> (comparison to PCS)
PCS	254	0.60 (0.04)	
MCS	254	0.57 (0.03)	.56
FACT trial outcome index	252	0.60 (0.03)	.99
FACT total	251	0.59 (0.04)	.94
BMT subscale	252	0.60 (0.03)	.88
FACT-BMT total	251	0.60 (0.03)	.90

AUC indicates area under the ROC curve; ROC, receiver operating characteristic; PCS, physical component score; MCS, mental component score; and BMT, BM transplantation.

HCT is one of recovery and return to normal functioning. Second, we have further demonstrated the magnitude of impairment in QOL by comparison of chronic GVHD cohort subjects' mean scores to those of age- and sex-matched US population normative data; these findings complement previously reported data.³⁵ Those affected by chronic GVHD had significant impairment in QOL across multiple domains. As well, we have for the first time examined the magnitude of impairment in QOL in chronic GVHD affected individuals in reference to that previously reported in the setting of other chronic health conditions. This frames the clinical relevance of the impairment observed in the context of the NIH severity staging, and demonstrates the marked impairment in physical functioning (PCS) but relatively preserved mental health domain (MCS) in chronic GVHD-affected individuals. A potentially important exception to this was the marked impairment in MCS in those with severe chronic GVHD, which rivaled that of depression.

We have also examined the discriminative accuracy of competing QOL instruments in this analysis, with the intention of determining which is most useful in evaluating the status of chronic GVHD severity. Using an extension of ROC methods, we found no significant differences between the SF-36 and FACT-BMT in discrimination of chronic GVHD severity. This cross-sectional analysis therefore does not support the superiority of one instrument over the other, and does not assist in the selection of QOL instrument for the purpose of investigation or clinical practice in chronic GVHD. We conclude that, while physical components of self-reported QOL are lower on average for patients with more severe cGVHD, the extent of impairment and symptom burden represented by cGVHD severity are not solely captured by differences in QOL. Future analyses will evaluate sensitivity to change and may help identify the better instrument to use in this population.

Several points are worthy to note as potential limitations to these findings. First, the power to characterize mild chronic GVHD category is limited by the small sample size. Small, but meaningful, differences (eg, differences in mean scores across the mild and moderate chronic GVHD categories) may not have been detected. Closer evaluation of the mild group awaits more patient data. Next, the composition of certain sociodemographic and transplant variables are notable. In regard to patient sociodemographic variables, cohort members were disproportionately white, non-Hispanic (mild 94%, moderate 87%, severe 90% of totals), thus limiting the

ability to generalize to other ethnic groups. Stem cell source used in HCT was near-uniformly peripheral blood mobilized stem cells (mild 81%, moderate 90%, and severe 90% of totals), but this reflects current US practice.

Finally, there are several relevant future directions moving beyond this analysis. Longitudinal assessment of QOL is ongoing and will provide more complete information on the duration of impairment, and trajectory of recovery or worsening. These data will permit analysis concerning the impact of resolved chronic GVHD on QOL; 2 prior studies have come to divergent conclusions regarding the durable impact of resolved chronic GVHD on QOL, but have suffered from relatively small numbers of chronic GVHD-affected subjects and methodologic limitations including the assessment of chronic GVHD activity by retrospective medical record abstraction.^{6,7} Longitudinal data can be used to assess whether the SF-36 or FACT-BMT are sensitive to changes in chronic GVHD severity to know whether it is realistic to expect that QOL can be used to measure treatment response.

In summary, the NIH global severity scoring system is correlated with patient-reported QOL, particularly in the physical domains, as detected by both the SF-36 and the FACT-BMT. These deficits are quite profound relative to the general population, and comparable with other chronic immune-mediated disorders.

Acknowledgment

This work was supported by National Institutes of Health/National Cancer Institute grant CA 118953-03 (PI: S.J.L.).

Authorship

Contribution: J.P. proposed the study concept, analyzed data, and wrote the manuscript; B.K. and X.C. performed statistical analyses and contributed to writing the manuscript; N.M., D.J.W., S.P., C.C., D.J., J.P., S.A., and M.J. contributed to data analysis and critical review of the manuscript; and S.J.L. contributed to the development of study concept, data analysis, and writing of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Joseph Pidala, MD, MS, Blood and Marrow Transplantation, Moffitt Cancer Center, 12902 Magnolia Dr, FOB 3308, Tampa, FL 33612; e-mail: joseph.pidala@moffitt.org.

References

- Buchanan DR, O'Mara AM, Kelaghan JW, Sgambati M, McCaskill-Stevens W, Minasian L. Challenges and recommendations for advancing the state-of-the-science of quality of life assessment in symptom management trials. *Cancer*. 2007;110(7):1621-1628.
- Halyard MY, Ferrans CE. Quality-of-life assessment for routine oncology clinical practice. *J Support Oncol*. 2008;6(5):221-229,233.
- Lee SJ, Fairclough D, Parsons SK, et al. Recovery after stem-cell transplantation for hematologic diseases. *J Clin Oncol*. 2001;19(1):242-252.
- Pidala J, Anasetti C, Jim H. Quality of life after allogeneic hematopoietic cell transplantation. *Blood*. 2009;114(1):7-19.
- Chiodi S, Spinelli S, Ravera G, et al. Quality of life in 244 recipients of allogeneic bone marrow transplantation. *Br J Haematol*. 2000;110(3):614-619.
- Fraser CJ, Bhatia S, Ness K, et al. Impact of chronic graft-versus-host disease on the health status of hematopoietic cell transplantation survivors: a report from the Bone Marrow Transplant Survivor Study. *Blood*. 2006;108(8):2867-2873.
- Kiss TL, Abdoell M, Jamal N, Minden MD, Lipton JH, Messner HA. Long-term medical outcomes and quality-of-life assessment of patients with chronic myeloid leukemia followed at least 10 years after allogeneic bone marrow transplantation. *J Clin Oncol*. 2002;20(9):2334-2343.
- Lee SJ, Kim HT, Ho VT, et al. Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant*. 2006;38(4):305-310.
- Syrjala KL, Chapko MK, Vitaliano PP, Cummings C, Sullivan KM. Recovery after allogeneic marrow transplantation: prospective study of predictors of long-term physical and psychosocial functioning. *Bone Marrow Transplant*. 1993;11(4):319-327.
- Worel N, Biener D, Kalhs P, et al. Long-term outcome and quality of life of patients who are alive and in complete remission more than two years after allogeneic and syngeneic stem cell transplantation. *Bone Marrow Transplant*. 2002;30(9):619-626.
- Duell T, van Lint MT, Ljungman P, et al. Health and functional status of long-term survivors of bone marrow transplantation. EBMT Working Party on Late Effects and EULEP Study Group on Late Effects. European Group for Blood and Marrow Transplantation. *Ann Intern Med*. 1997;126(3):184-192.
- Lee SJ, Flowers ME. Recognizing and managing chronic graft-versus-host disease. *Hematology Am Soc Hematol Educ Program*. 2008:134-141.
- Stewart BL, Storer B, Storek J, et al. Duration of immunosuppressive treatment for chronic graft-versus-host disease. *Blood*. 2004;104(12):3501-3506.
- 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.
15. Couriel D, Carpenter PA, Cutler C, et al. Ancillary therapy and supportive care of chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease. V. Ancillary Therapy and Supportive Care Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(4):375-396.
 16. Pavletic SZ, Martin P, Lee SJ, et al. Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease. IV. Response Criteria Working Group report. *Biol Blood Marrow Transplant*. 2006;12(3):252-266.
 17. Schultz KR, Miklos DB, Fowler D, et al. Toward biomarkers for chronic graft-versus-host disease: National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease. III. Biomarker Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(2):126-137.
 18. Shulman HM, Kleiner D, Lee SJ, et al. Histopathologic diagnosis of chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease. II. Pathology Working Group Report. *Biol Blood Marrow Transplant*. 2006;12(1):31-47.
 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. 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.
 21. McQuellon RP, Russell GB, Cella DF, et al. Quality of life measurement in bone marrow transplantation: development of the Functional Assessment of Cancer Therapy-Bone Marrow Transplant (FACT-BMT) scale. *Bone Marrow Transplant*. 1997;19(4):357-368.
 22. Functional Assessment of Chronic Illness Therapy. www.facit.org. Accessed August 2010.
 23. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31(3):247-263.
 24. Ware JE, Kosinski M, Keller SD. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston, MA: The Health Institute; 1994.
 25. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston, MA: The Health Institute; 1993.
 26. Ware JE KM, Dewey J. *How to Score Version 2 of the SF-36 Health Survey*. Lincoln, RI: QualityMetric Inc; 2000.
 27. Quality Metric: SF Health Surveys. www.qualitymetric.com. Accessed August 2010.
 28. Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol*. 2005;12(9):1198-1204.
 29. Nguyen P. nonbinROC: software for evaluating diagnostic accuracies with non-binary gold standards. *J Stat Softw*. 2007;21(10):1-10.
 30. Alarcon GS, McGwin G Jr, Uribe A, et al. Systemic lupus erythematosus in a multiethnic lupus cohort (LUMINA). XVII. Predictors of self-reported health-related quality of life early in the disease course. *Arthritis Rheum*. 2004;51(3):465-474.
 31. Del Rosso A, Boldrini M, D'Agostino D, et al. Health-related quality of life in systemic sclerosis as measured by the Short Form 36: relationship with clinical and biologic markers. *Arthritis Rheum*. 2004;51(3):475-481.
 32. Khanna D, Clements PJ, Furst DE, et al. Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis: results from the Scleroderma Lung Study. *Arthritis Rheum*. 2005;52(2):592-600.
 33. Mahler DA, Mackowiak JI. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with COPD. *Chest*. 1995;107(6):1585-1589.
 34. Pittock SJ, Mayr WT, McClelland RL, et al. Quality of life is favorable for most patients with multiple sclerosis: a population-based cohort study. *Arch Neurol*. 2004;61(5):679-686.
 35. Mitchell SA, Leidy NK, Mooney KH, et al. Determinants of functional performance in long-term survivors of allogeneic hematopoietic stem cell transplantation with chronic graft-versus-host disease (cGVHD). *Bone Marrow Transplant*. 2010; 45(4):762-769.