

Sensitivity of changes in chronic graft-versus-host disease activity to changes in patient-reported quality of life: results from the Chronic Graft-versus-Host Disease Consortium

Joseph Pidala,¹ Brenda F. Kurland,² Xiaoyu Chai,² Georgia Vogelsang,³ Daniel J. Weisdorf,⁴ Steven Pavletic,⁵ Corey Cutler,⁶ Navneet Majhail,⁴ and Stephanie J. Lee²

¹Moffitt Cancer Center, Tampa, FL; ²Fred Hutchinson Cancer Research Center, Seattle, WA; ³Johns Hopkins University, Baltimore, MD; ⁴University of Minnesota, Minneapolis, MN; ⁵National Cancer Institute, Bethesda, MD and ⁶Dana Farber Cancer Institute, Boston, MA, USA

ABSTRACT

Background

The 2005 National Institute of Health Chronic Graft-versus-Host Disease Consensus Conference recommended collection of patient-reported outcomes in clinical trials on chronic graft-versus-host disease. We assessed whether changes in chronic graft-versus-host disease severity, determined using National Institute of Health criteria, clinicians' assessment or patients' self-evaluation, correlated with patient-reported quality of life as measured by the Short Form-36 and Functional Assessment of Cancer Therapy – Bone Marrow Transplant (FACT-BMT) instruments.

Design and Methods

Three-hundred and thirty-six adult patients (median age 52 years; range, 19 – 79) with chronic graft-versus-host disease from six transplant centers contributed baseline and follow-up data (from 936 visits overall).

Results

While the majority of the patients had stable chronic graft-versus-host disease, improvement or worsening was noted in approximately 40% of follow-up visits. Multivariable analysis demonstrated no association between change in chronic graft-versus-host disease severity evaluated by National Institute of Health criteria and change in quality of life, while clinician-reported changes in severity were associated with changes in some quality of life measures. Patient-reported changes in the severity of chronic graft-versus-host disease were associated with changes in all quality of life measures. Comparison of the Short Form-36 and the FACT-BMT suggested that the data collected in the Functional Assessment of Cancer Therapy – General (FACT-G) core survey are sufficient without the need for the Short Form-36 or the FACT-BMT subscale.

Conclusions

We conclude that serial National Institute of Health and clinician-reported chronic graft-versus-host disease severity assessments cannot substitute for patient-reported outcomes in clinical trials. Collection of just the FACT-G instead of the Short Form-36 and the full FACT-BMT will decrease respondent burden without compromising quality of life assessment.

Key words: chronic graft versus host disease, quality of life, sensitivity to change.

Citation: Pidala J, Kurland BF, Chai X, Vogelsang G, Weisdorf DJ, Pavletic S, Cutler C, Majhail N, and Lee SJ. Sensitivity of changes in chronic graft-versus-host disease activity to changes in patient-reported quality of life: results from the Chronic Graft-versus-Host Disease Consortium. *Haematologica* 2011;96(10):1528-1535. doi:10.3324/haematol.2011.046367

©2011 Ferrata Storti Foundation. This is an open-access paper.

Manuscript received on April 19, 2011. Revised version arrived on June 8, 2011. Manuscript accepted on June 10, 2011

Correspondence:
Joseph Pidala, MD, MS, Blood and Marrow Transplantation, Moffitt Cancer Center 12902 Magnolia Drive, FOB 3308 Tampa, FL 33612 USA.
Phone: international +813.7452556.
Fax: international +813.4498248.
E-mail: joseph.pidala@moffitt.org

Presented in abstract form at the ASBMT/CIBMTR meeting, Honolulu, 2/2011

Introduction

Chronic graft-versus-host disease (GVHD) is an immune-mediated disorder which is associated with late hematopoietic cell transplantation (HCT)-related morbidity, mortality, prolonged duration of immune suppressive therapy, and impaired quality of life (QOL).¹⁻⁴ While the anticipated trajectory following HCT for patient-reported QOL is one of recovery and return to pre-HCT functioning, previous studies have demonstrated that those affected by chronic GVHD suffer ongoing impairment in multiple domains of QOL.⁵⁻¹⁴ The global severity of chronic GVHD as determined by National Institute of Health (NIH) criteria is associated with QOL, and patients with moderate to severe chronic GVHD have significant impairment across multiple domains of QOL as shown by comparing their scores with population normative scores.¹⁵ Those with moderate to severe chronic GVHD have Short Form-36 (SF-36) physical component scores (PCS) similar to those reported for patients with allied immune-mediated disorders such as systemic sclerosis, systemic lupus erythematosus, and multiple sclerosis, and greater impairment compared to patients with common chronic medical conditions.¹⁵

However, these analyses have not considered clinician-assessed or patient-reported severity of chronic GVHD, and have not characterized the relevance of dynamic changes in chronic GVHD activity on patient-reported outcomes. To address these limitations, we performed an analysis of the association between changes in chronic GVHD severity as assessed by NIH global severity criteria, clinicians, and patients with changes in patient-reported QOL in a multicenter cohort of HCT recipients affected by chronic GVHD.

Design and Methods

Patients with chronic graft-versus-host disease

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 six sites (Fred Hutchinson Cancer Research Center, University of Minnesota, Dana-Farber Cancer Institute, Stanford University, Vanderbilt University, and the Medical College of Wisconsin). All subjects provided written informed consent to participation. Subjects included were allogeneic HCT recipients with chronic GVHD requiring systemic immunosuppressive therapy. Patients with classic chronic GVHD and those with overlap syndrome were eligible.¹⁶ Cases were classified as incident (enrollment less than 3 months after the diagnosis of chronic GVHD) or prevalent (enrollment 3 or more months but less than 3 years after the diagnosis of chronic GVHD). Exclusion criteria were primary disease relapse after HCT and inability to comply with study procedures. At enrollment and every 6 months thereafter (incident cases had an extra assessment 3 months after enrollment), clinicians and patients reported standardized information on chronic GVHD organ involvement, severity, and symptoms. The global severity of the chronic GVHD according to the NIH Consensus criteria was calculated from individual organ scores provided by clinicians using the NIH consensus scoring algorithm, resulting in a global grade of none, mild, moderate, or severe.¹⁶ Chronic GVHD severity was also independently assessed by both clinicians' and patients' report at each study visit using the qualitative categories of "none", "mild", "moderate" and "severe" without specific definitions. Medical charts were reviewed to collect objective medical data, laboratory and examination results, information on compli-

cations, and medication profiles.

All available longitudinal data on chronic GVHD severity and QOL were included in this analysis, from cohort inception through March, 2010, thus including both baseline enrollment data and data from all available follow-up assessments. Off study visits when a patient relapsed or died were excluded from this analysis since patient-reported information was rarely available.

Data collection forms, the ACCESS database structure, and SAS code for the NIH consensus severity algorithm are available upon request from the authors.

Quality of life instruments

The Functional Assessment of Cancer Therapy – Bone Marrow Transplant (FACT-BMT) and SF-36 questionnaires were completed by patients age 18 years or older. The FACT-BMT v4.0 is a 37-item self-report QOL questionnaire, which includes a 10-item BMT subscale. The instrument measures the effect of cancer therapy on domains including physical, functional, social/family, and emotional well-being, plus BMT-specific concerns. Individual core domain scores are summarized to give a Functional Assessment of Cancer Therapy – General (FACT-G) score based on 27 items. The FACT-G plus the BMT subscale yield a FACT-BMT score based on all 37 items. Additionally, a Functional Assessment of Cancer Therapy – Trial Outcome Index (FACT-TOI) score consists of the sum of physical and functional well-being subscales and the BMT subscale (24 items).¹⁷ We analyzed these three overlapping summary scores to see if we could identify a parsimonious battery that would capture the outcomes of interest with the minimum number of items.

The SF-36 v2 is a 36-item self-report questionnaire which assesses the following domains of QOL: physical functioning, role functioning-physical, bodily pain, general health, vitality, social functioning, role functioning-emotional, and mental health. Two summary scales from the SF-36 are the PCS and the mental component score (MCS).¹⁸⁻²⁰ The PCS and MCS are normalized so that a score of 50 represents the average for the 1998 United States general population with a standard deviation of 10. In validation studies, the correlation between the SF-36v1.0 and the SF-12 was 91%.²¹

Statistical methods

Standard algorithms were used to compute total and subscale scores for the FACT-BMT¹⁷ and SF-36 instruments.^{19,20} Chronic GVHD severity was separately determined according to the three severity grading schemes (NIH consensus criteria, clinician-assessed severity, and patient-reported chronic GVHD severity) for each of the study visits.

Change in chronic GVHD severity between sequential visits was calculated for each grading scheme. The continuous change scale was also categorized into stable (0), improvement (negative values), and worsening (positive values) chronic GVHD severity, so that asymmetry between QOL sensitivity to improvement *versus* worsening could be detected in regression models. Agreement regarding change in severity for the three grading schemes was calculated using weighted kappa for ordinal measures with Fleiss-Cohen weights.²² Missing QOL data occurred when an office visit was completed, but the patient did not complete the QOL questionnaires at the visit, or did not return them by mail afterwards. A dependence of missing QOL data on health status could result in bias in assessments of associations between change in QOL and change in severity. We, therefore, summarized patterns for missing QOL data with respect to patients' socio-demographic and chronic GVHD involvement and severity characteristics. For baseline characteristics, statistical comparisons between groups were made with the two-sample t-test of independence for continuous variables, and Fisher's exact test for categorical variables. Additionally,

patterns of missing data by chronic GVHD severity and organ involvement were compared using generalized estimating equations to adjust for multiple observations per patient. If observed covariates were related to missing data, including these variables as predictors in the linear mixed models could assist in implicit imputation of missing QOL data and prevent bias in assessing the relationship between change in QOL and change in chronic GVHD severity.²³

Multivariable models were constructed for the primary analysis measuring the association between changes in chronic GVHD severity and changes in patient-reported QOL, using linear mixed models to account for within-patient correlation and data missing at random.²³ Separate models were fitted for each of the QOL instruments/domains of interest (SF-36 PCS, SF-36 MCS, FACT-TOI, FACT-G, FACT-BMT). Patients' socio-demographic, disease, and transplantation covariates considered in the models included the following: age, gender, education, income level, time from HCT to enrollment, disease stage, donor relation, conditioning regimen intensity, incident *versus* prevalent chronic GVHD, and transplant center. Effects of selected covariates were held constant when exploring relationships between change in patient-reported QOL and stability, worsening, or improvement in chronic GVHD severity.

A final, complementary analysis examined concordance between changes in each severity grading scheme and clinically meaningful change in QOL measures, using weighted kappa for ordinal measures with Fleiss-Cohen weights.²² The threshold for improvement or worsening in QOL was one-half of the population standard deviation (5 points for SF-36, or 7 points for FACT).

Statistical analyses were conducted using SAS/STAT software, version 9.2 (SAS Institute, Inc., Cary, NC, USA) and R version 2.9.2 (R Foundation for Statistical Computing, Vienna, Austria). Recognizing the multiple comparisons involved in comparing several QOL measures and pair-wise comparisons among GVHD severity levels, a type I error was controlled by considering a *P* value of 0.02 or lower as statistically significant, and by looking for consistency of results across related constructs.

Results

Characteristics of the cohort of patients with chronic graft-versus-host disease

This analysis was conducted on longitudinal data from 336 patients, including data from enrollment visits and 600 follow-up visits through March, 2010. The analysis, therefore, encompasses chronic GVHD severity and QOL data from a total of 936 visits. Baseline socio-demographic, disease, and transplantation characteristics, as well as baseline SF-36 and FACT-BMT scores are summarized in Table 1. The cohort included nearly equal numbers of incident (53%) and prevalent (47%) cases of chronic GVHD. Both male (60%) and female (40%) HCT recipients were well-represented, as were recipients of myeloablative (57%) and non-myeloablative (43%) transplants. Skin, mouth, eye, and lung symptoms were present in at least half of visits; genital involvement was least commonly reported, in only 13% of visits. Average SF-36 subscale scores were generally somewhat below the population norm of 50 points, with the exception of the mental health scale. The lowest average values were found for the role-functioning physical scale (39.5 points; range, 17.7 – 56.9) and the PCS (40.6 points; range 6.2 – 61.1).

A total of 142 patients had 203 visits at which SF-36

and/or FACT-BMT data were missing. Of these, 97 patients had data missing for one visit, 30 for two visits, 14 for three visits, and one for four visits. The most common scenario was lacking data for both the SF-36 and FACT-BMT for a visit, as compared to missing just one instrument, suggesting that patients failed to complete the entire patient-reported battery of instruments. Table 2 shows comparisons of patients with any missing data (n=142) to patients with no missing QOL assessments (n=194). Patients who completed all QOL assessments had a higher educational level than patients with some missing QOL data (*P*=0.01) but were otherwise similar in gender and case type at enrollment (incident *versus* prevalent, overlap *versus* classic chronic GVHD).

Visits for which QOL data were missing (Table 3) tended to occur at a longer time since enrollment (median 6 *versus* 4 months, *P*=0.01 for comparison of means) and after a longer interval between visits. The mean time interval between visits for incident cases was 4.2 months, while it was 6.1 months for prevalent cases. After accounting for patient case type, longer time span between visits was moderately associated with a higher chance of missing SF36 and/or FACT-BMT data (*P*=0.04). For example, the estimated odds of missing data for at least one instrument was 57% higher for visits occurring after an interval of more than 6 months, comparing to visits occurring after an interval from 3 – 6 months (95% CI: 2% higher – 143% higher). Organ involvement and overall chronic GVHD severity as assessed by NIH criteria, clinician assessment or patient assessment were similar (Table 3).

Change in chronic graft-versus-host disease severity on serial assessment

Change in chronic GVHD severity was examined by comparing severity at each visit to severity at the previous visit. In most cases the serial assessments showed that the severity of the patients' chronic GVHD remained stable (NIH 63%, clinician 63%, or patient 58%). However, follow-up visits demonstrated improvement (NIH 19%, clinician 22%, patient 24%) or worsening (NIH 18%, clinician 15%, patient 18%) in others. The proportion of patients with improvement in chronic GVHD severity by any measure was highest between the first and second visits, perhaps reflecting response to treatment initiated at the time of study entry but remained close to 40% even at subsequent visits (*data not shown*). Change in the NIH global severity score was largely driven by change in severity in skin (Pearson's correlation coefficient *r* = 0.32), eye (*r* = 0.30), and lung (*r* = 0.49) involvement. Change in severity in other affected organs was correlated to a lesser extent with change in the NIH global severity score (mouth, *r*=0.21; gastrointestinal tract, *r*=0.08; liver, *r*=0.14; joint, *r*=0.04; genital, *r*=0.14).

Although the prevalences of stability, improvement, and worsening were similar according to NIH criteria, clinicians' assessment and patients' ratings, the agreements were not strong. As shown in Figure 1, weighted kappa values for agreement between change in pairs of measures were low, ranging from 0.06 to 0.23. For example, Figure 1B shows the agreement between change in severity according to NIH criteria and according to patients' report. Worsening according to NIH criteria and improvement according to patients' report (upper left corner of Figure 1B) was roughly as common as worsening according to both criteria (upper right corner). These discrepancies between clinician-assessed and

Table 1. Characteristics and QOL instrument scores for the cohort of patients with chronic GVHD.

| Characteristics | N | n (%) | Mean | SE | Median | Min | Max |
|---|-----|-----------|------|-----|--------|------|------|
| Patient level (N = 336) | | | | | | | |
| Patient age at registration (years) | 336 | | 50.2 | 0.7 | 52.0 | 19.0 | 79.0 |
| Months from transplant to enrollment | 336 | | 15.0 | 0.5 | 12.3 | 2.9 | 39.2 |
| Patient gender | 336 | | | | | | |
| Female | | 135 (40%) | | | | | |
| Male | | 201 (60%) | | | | | |
| Site | 336 | | | | | | |
| Fred Hutchinson Cancer Research Center | | 168 (50%) | | | | | |
| University of Minnesota Dana-Faber Cancer Institute | | 42 (13%) | | | | | |
| Stanford University Medical Center | | 41 (12%) | | | | | |
| Vanderbilt University Medical Center | | 51 (15%) | | | | | |
| Medical College of Wisconsin | | 28 (8%) | | | | | |
| of Wisconsin | | 6 (2%) | | | | | |
| Disease status | 336 | | | | | | |
| Early | | 111 (33%) | | | | | |
| Intermediate | | 153 (46%) | | | | | |
| Advanced | | 72 (21%) | | | | | |
| Transplant type | 336 | | | | | | |
| Myeloablative | | 191 (57%) | | | | | |
| Non-myeloablative | | 145 (43%) | | | | | |
| Case type | 336 | | | | | | |
| Incident | | 178 (53%) | | | | | |
| Prevalent | | 158 (47%) | | | | | |
| Donor match | 336 | | | | | | |
| HLA matched relative | | 154 (46%) | | | | | |
| HLA mismatched relative | | 7 (2%) | | | | | |
| Unrelated donor | | 175 (52%) | | | | | |
| Marital status | 291 | | | | | | |
| Married/living with partner | | 231 (79%) | | | | | |
| Single, never married | | 21 (7%) | | | | | |
| Divorced, separated | | 29 (10%) | | | | | |
| Widowed | | 9 (3%) | | | | | |
| Other | | 1 (1%) | | | | | |
| Highest education | 291 | | | | | | |
| Some high school | | 4 (1%) | | | | | |
| High school graduate | | 48 (16%) | | | | | |
| Some college | | 95 (33%) | | | | | |
| College graduate | | 78 (27%) | | | | | |
| Post-graduate degree | | 66 (23%) | | | | | |
| Annual family income | 271 | | | | | | |
| Under \$15,000 | | 12 (4%) | | | | | |
| \$15,000 - \$24,999 | | 14 (5%) | | | | | |
| \$25,000 - \$49,999 | | 38 (14%) | | | | | |
| \$50,000 - \$74,999 | | 61 (23%) | | | | | |
| \$75,000 - \$99,999 | | 48 (18%) | | | | | |
| \$100,000 or above | | 98 (36%) | | | | | |
| Visit level (N = 936) | | | | | | | |
| Current GVHD classification | 852 | | | | | | |
| Classic chronic GVHD | | 655 (77%) | | | | | |
| Overlap acute and chronic GVHD | | 197 (23%) | | | | | |
| Skin involvement | 896 | 507 (57%) | | | | | |
| Eye involvement | 896 | 495 (55%) | | | | | |
| Mouth involvement | 895 | 479 (54%) | | | | | |

| | | | | | | | |
|---|-----|-----------|------|-----|------|------|------|
| Liver involvement | 897 | 350 (39%) | | | | | |
| Joint involvement | 896 | 286 (32%) | | | | | |
| Lung involvement | 896 | 459 (51%) | | | | | |
| Gastrointestinal involvement | 895 | 194 (22%) | | | | | |
| Genital involvement | 773 | 100 (13%) | | | | | |
| Months elapsed between visits (excludes baseline) | 600 | | 5.1 | 1.8 | 5.4 | 1.4 | 11.3 |
| SF36-physical functioning | 762 | | 42.4 | 0.4 | 44.4 | 14.9 | 57.0 |
| SF36-role physical | 767 | | 39.5 | 0.4 | 39.7 | 17.7 | 56.9 |
| SF36-bodily pain | 765 | | 46.3 | 0.4 | 46.1 | 19.9 | 62.1 |
| SF36-general health | 763 | | 42.2 | 0.4 | 41.0 | 16.2 | 63.9 |
| SF36-vitality | 766 | | 47.0 | 0.4 | 49.0 | 20.9 | 70.8 |
| SF36-social functioning | 769 | | 42.3 | 0.4 | 45.9 | 13.2 | 56.8 |
| SF36-role emotional | 763 | | 46.0 | 0.4 | 48.1 | 9.2 | 55.9 |
| SF36-mental health | 766 | | 50.4 | 0.4 | 52.8 | 7.8 | 64.1 |
| SF36-physical component scale | 747 | | 40.6 | 0.4 | 41.5 | 6.2 | 61.1 |
| SF36-mental component scale | 747 | | 49.3 | 0.4 | 52.0 | 6.9 | 69.0 |
| FACT-physical | 765 | | 21.3 | 0.2 | 22.0 | 4.0 | 28.0 |
| FACT-social/family | 768 | | 22.4 | 0.2 | 23.0 | 0.0 | 28.0 |
| FACT-emotional | 766 | | 19.3 | 0.1 | 20.0 | 2.0 | 24.0 |
| FACT-functional | 761 | | 18.0 | 0.2 | 18.0 | 3.0 | 28.0 |
| FACT-BMT subscale | 760 | | 27.7 | 0.2 | 28.0 | 11 | 40.0 |
| FACT-BMT TOI | 757 | | 67.0 | 0.6 | 68.0 | 21 | 96.0 |
| FACT-G | 755 | | 81.0 | 0.6 | 83.0 | 23 | 108 |
| FACT-BMT total score | 754 | | 109 | 0.8 | 110 | 36 | 148 |

SE: standard error; SF36: short form 36; FACT: functional assessment of cancer therapy.

patient-assessed changes in GVHD severity could be associated with physical functioning level as measured by the SF-36 PCS, or with specific organ manifestations (yes/no for skin, mouth, eye, gastrointestinal tract, liver, joint, genital, or lung involvement). Generalized estimating equations were used to model agreement in the three-level classification of severity change; none of these cofactors was related to agreement of clinicians' and patients' assessments ($P \geq 0.1$).

Associations between change in chronic graft-versus-host disease severity and change in quality of life

Multivariable analysis considered patients' socio-demographic, disease, and transplantation covariates (age, gender, education, income level, time from HCT to enrollment, disease stage, donor relation, conditioning regimen intensity, incident versus prevalent chronic GVHD, and transplant center) as predictors of QOL, in addition to three levels of change in chronic GVHD severity ratings (stable, improved, worsened). Separate models were constructed for each of the QOL outcomes (SF-36 PCS, SF-36 MCS, FACT-TOI, FACT-G, and FACT-BMT). For each model, the change in chronic GVHD severity was also analyzed as a continuous predictor (to use the full range of change scores) and as a categorical predictor (to allow for different magnitudes of effects on QOL change for improvement and worsening). In univariable models, statistically significant associations were not found between any socio-demographic, disease, or transplantation covariates and change in PCS, FACT-

TOI, or FACT-BMT. For change in FACT-G, a positive association with education level persisted in multiple exploratory models. Since education and months since enrollment were associated with missing data (Tables 2 and 3), multivariable models adjusted for education level and months since enrollment for all QOL outcomes.

Table 4 shows the multivariable results, with separate models for chronic GVHD change as a continuous predictor of change in QOL (center columns), and as a categorical predictor (right-hand columns). Change in GVHD severity assessed using NIH criteria was not associated with change in any of the QOL outcomes. This conclusion held true for both the studied summary scores (SF-36 PCS, SF-36 MCS, FACT-TOI, FACT-G, and FACT-BMT), as well as detailed investigation of individual domain scores for each (*data not shown*). Change in clinician-reported severity of GVHD was associated with changes in QOL assessed by FACT-G and FACT-BMT. For the FACT-G, FACT-BMT and SF-36 MCS, as well as the individual domain scores which contribute to these summary scores, clinician-assessed improvement (*versus* stability) in chronic GVHD severity was more consistently associated with change in QOL than worsening (*versus* stability). Patient-reported change in severity of GVHD

was sensitive to changes in all QOL outcomes, including both the summary scores, and individual domain scores, and for both improvement and worsening in chronic GVHD severity.

Association between change in chronic graft-versus-host disease severity and clinically meaningful change in quality of life

To complement the analysis in Table 4, we next examined the relationship between change in chronic GVHD severity according to each severity grading scheme and clinically meaningful change in QOL. We defined clinically meaningful change according to one-half of the population standard deviation (5 points for SF-36, or 7 points for FACT). An overall small proportion of the patients at each of the chronic GVHD cohort visits had clinically meaningful improvement or worsening (approximately 20% each, with the remaining stable) on follow-up across each of the examined QOL domains (SF-36 PCS, SF-36 MCS, FACT-TOI, FACT-G, and FACT-BMT, Table 5). Weighted kappa statistics were calculated to measure the degree of concordance and showed low concordance between clinically meaningful change in QOL and change in the severity of chronic

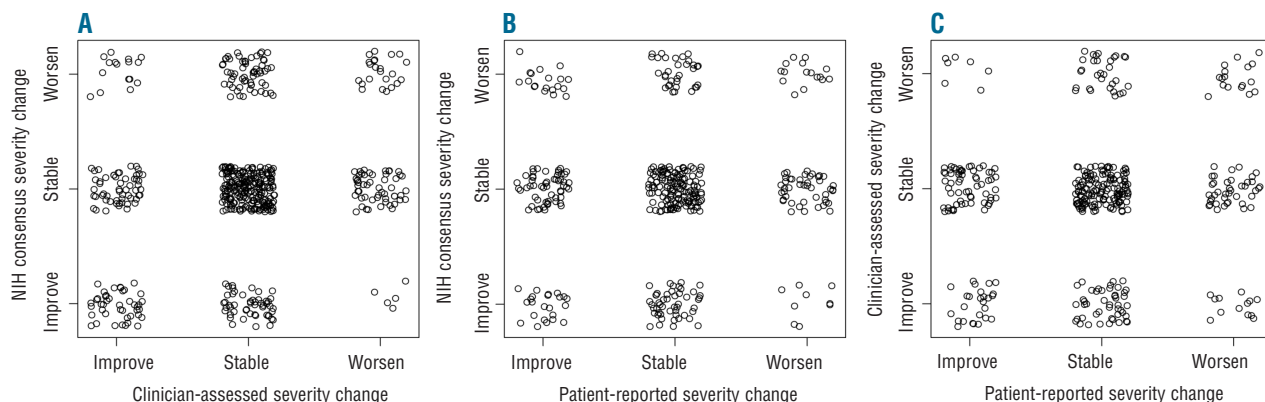


Figure 1. Agreement between the NIH, clinician-, and patient-determined chronic GVHD severity changes. *Jittered* points were displayed to differentiate tied data in each category. Fleiss-Cohen weighted kappa was used to measure agreement. (A) NIH consensus and clinicians' assessment ($\kappa=0.23$); (B) NIH consensus and patients' assessment ($\kappa=0.06$). (C) Clinicians' assessment and patients' assessment ($\kappa=0.14$).

Table 2. Patient-level description of characteristics related to missing QOL data.

| | Patients with no missing QOL (N = 194) | | | | Patients lacking SF36 and/or FACT-BMT date (N = 142) | | | | P value* | | |
|---------------------------|--|-----|--------|-----|--|----------|-----|--------|----------|-----|------|
| | n (%) | n | Median | Min | Max | n (%) | n | Median | | Min | Max |
| Age at enrollment (years) | | 194 | 52.5 | 19 | 74 | | 142 | 52 | 19 | 79 | 0.40 |
| Gender | | 194 | | | | | 142 | | | | 0.57 |
| Male | 119 (61%) | | | | | 82 (58%) | | | | | |
| Female | 75 (39%) | | | | | 60 (42%) | | | | | |
| Education level | | 193 | | | | | 98 | | | | 0.01 |
| Some high school | 3 (2%) | | | | | 1 (1%) | | | | | |
| High school graduate | 25 (13%) | | | | | 23 (23%) | | | | | |
| Some college | 59 (30%) | | | | | 36 (37%) | | | | | |
| College graduate | 52 (27%) | | | | | 26 (27%) | | | | | |
| Post graduate degree | 54 (28%) | | | | | 12 (12%) | | | | | |
| Case type | | 194 | | | | | 142 | | | | 0.22 |
| Incident | 97 (50%) | | | | | 81 (57%) | | | | | |
| Prevalent | 97 (50%) | | | | | 61 (43%) | | | | | |

*P values are from two-sample t-test of independence for continuous variables, and Fisher's exact test for categorical variables. QOL, quality of life; SF36, Short Form 36; FACTBMT, functional Assessment of Cancer Therapy – Bone Marrow Transplant Subscale.

GVHD. We did, however, observe a trend in which agreement was greater between clinically meaningful change in QOL and change in patient-reported chronic GVHD severity, as compared to that observed with change in chronic GVHD severity assessed either using NIH criteria or by clinicians (Table 5).

Discussion

Patient-reported QOL represents a comprehensive assessment of patients' perceived functional ability and overall well-being, and is cited by both patients and providers as an important goal of treatment. Measurable improvements in QOL may also be recognized by the Food and Drug Administration as evidence of "clinical benefit", a necessary component for regulatory approval.²⁴ The collection of patient-reported data was, therefore, recommended by the NIH Consensus Conference as a valuable outcome measure in clinical practice and therapeutic trials.

We made several important findings. First, change in NIH-assessed global chronic GVHD severity over a 6-month average timeframe was not associated with change in any of the studied QOL domains. In contrast, changes in

clinician-assessed chronic GVHD were associated with changes in the FACT-G and FACT-BMT scores of QOL, primarily when chronic GVHD became less severe. Change in patient-reported chronic GVHD severity, either improvement or worsening, was strongly correlated with change in QOL, demonstrating significant associations on multivariable modeling with all of the studied QOL domains of both the SF-36 and FACT-BMT. While this observation is somewhat intuitive, it confirms that if one wants to know whether patients perceive a change in the severity of their chronic GVHD or whether their QOL has changed, the only way to ascertain this information in clinical trials or in clinical practice is to ask patients directly. One cannot rely on clinician-reported chronic GVHD organ scores or clinician-reported severity of GVHD to extrapolate changes in QOL.

We also observed that patients demonstrated clinically relevant improvement or worsening in QOL (change greater than 0.5 times the population standard deviation) compared to that at the prior visit in approximately 40% of follow-up visits. These clinically meaningful changes showed the greatest correlation with change in patient-reported chronic GVHD severity, and little to no correlation with NIH-calculated or clinician-reported changes in

Table 3. Visit-level description of characteristics related to missing QOL data.

| | Visits with no missing QOL (N = 733) | | | | Visits lacking SF36 or FACT-BMT date (N = 203) | | | | P value** | | |
|--|---|-----|--------|-----|---|-----------|-----|--------|-----------|------|------|
| | n (%) | n | Median | Min | Max | n (%) | n | Median | | Min | Max |
| Months from enrollment to current visit | | 733 | 4 | 0 | 29 | | 203 | 6 | 0 | 24 | 0.01 |
| Months from prior visit to current visit (excludes baseline) | | 456 | 5.3 | 1.5 | 9.7 | | 144 | 5.8 | 1.4 | 11.3 | 0.06 |
| NIH consensus severity | | 727 | | | | | 169 | | | | 0.33 |
| None | 11 (2%) | | | | | 3 (2%) | | | | | |
| Mild | 93 (13%) | | | | | 17 (10%) | | | | | |
| Moderate | 423 (58%) | | | | | 97 (57%) | | | | | |
| Severe | 200 (27%) | | | | | 52 (31%) | | | | | |
| Clinician-assessed severity | | 723 | | | | | 169 | | | | 0.66 |
| None | 28 (4%) | | | | | 8 (5%) | | | | | |
| Mild | 405 (56%) | | | | | 98 (58%) | | | | | |
| Moderate | 242 (33%) | | | | | 48 (28%) | | | | | |
| Severe | 48 (7%) | | | | | 15 (9%) | | | | | |
| Patient-reported severity | | 712 | | | | | 34 | | | | 0.56 |
| None | 51 (7%) | | | | | 2 (6%) | | | | | |
| Mild | 395 (56%) | | | | | 18 (53%) | | | | | |
| Moderate | 228 (32%) | | | | | 11 (32%) | | | | | |
| Severe | 38 (5%) | | | | | 3 (9%) | | | | | |
| Current chronic GVHD type | | 690 | | | | | 162 | | | | 0.34 |
| Overlap | 164 (24%) | | | | | 33 (20%) | | | | | |
| Classic | 526 (76%) | | | | | 129 (80%) | | | | | |
| Organ involved | | | | | | | | | | | |
| Skin | 410 (56%) | 727 | | | | 97 (57%) | 169 | | | | 0.85 |
| Mouth | 378 (52%) | 727 | | | | 101 (60%) | 168 | | | | 0.07 |
| Eye | 411 (57%) | 727 | | | | 84 (50%) | 169 | | | | 0.92 |
| Gastrointestinal | 157 (22%) | 726 | | | | 37 (22%) | 169 | | | | 0.97 |
| Liver | 281 (39%) | 727 | | | | 69 (41%) | 170 | | | | 0.85 |
| Joint | 230 (32%) | 727 | | | | 56 (33%) | 169 | | | | 0.89 |
| Genital | 82 (13%) | 628 | | | | 18 (12%) | 145 | | | | 0.30 |
| Lung | 370 (51%) | 727 | | | | 89 (53%) | 169 | | | | 0.66 |

** P values are from models fitted using generalized estimating equations to adjust for multiple observations per patient. QOL: quality of life; SF36: Short Form-36; FACT-BMT: Functional Assessment of Cancer Therapy – Bone Marrow Transplant subscale; NIH: National Institutes of Health; GVHD: graft-versus-host disease.

Table 4. Results of multivariable analyses examining relationships between changes in chronic GVHD severity and changes in patient-reported QOL, holding education level and months since enrollment constant. Linear mixed models with random patient effects were fitted, with separate models for each of the QOL outcomes (SF-36 PCS and SF36-MCS; FACT-TOI, FACT-G, and FACT-BMT). Center columns describe models with chronic GVHD change as a continuous predictor of change in QOL, and right-hand columns model chronic GVHD change as a categorical predictor. Figures in bold $P \leq 0.02$.

| Outcome | Covariates* | Estimate | P value | 95% lower CI | 95% upper CI | Contrast | Estimate | P value | 95% lower CI | 95% upper CI |
|------------------------|------------------------------------|--------------|------------------|--------------|--------------|--------------------|--------------|------------------|--------------|--------------|
| SF36-PCS change | | | | | | | | | | |
| | NIH consensus severity change | 0.08 | 0.881 | -1.03 | 1.20 | Improve vs. Stable | -0.79 | 0.430 | -2.75 | 1.18 |
| | | | | | | Worsen vs. Stable | -0.73 | 0.480 | -2.77 | 1.31 |
| | Clinician-assessed severity change | -0.14 | 0.815 | -1.34 | 1.06 | Improve vs. Stable | -0.75 | 0.436 | -2.66 | 1.15 |
| | | | | | | Worsen vs. Stable | -1.12 | 0.329 | -3.38 | 1.14 |
| | Patient-reported severity change | -2.77 | <0.001 | -3.83 | -1.70 | Improve vs. Stable | 2.76 | 0.004 | 0.91 | 4.61 |
| | | | | | | Worsen vs. Stable | -3.28 | 0.002 | -5.36 | -1.21 |
| SF36-MCS change | | | | | | | | | | |
| | NIH consensus severity change | -0.56 | 0.382 | -1.83 | 0.70 | Improve vs. Stable | 1.78 | 0.115 | -0.44 | 4.00 |
| | | | | | | Worsen vs. Stable | 0.97 | 0.407 | -1.34 | 3.28 |
| | Clinician-assessed severity change | -1.56 | 0.025 | -2.91 | -0.20 | Improve vs. Stable | 3.15 | 0.004 | 1.00 | 5.30 |
| | | | | | | Worsen vs. Stable | 0.84 | 0.518 | -1.71 | 3.38 |
| | Patient-reported severity change | -2.74 | <0.001 | -3.94 | -1.55 | Improve vs. Stable | 2.54 | 0.017 | 0.45 | 4.62 |
| | | | | | | Worsen vs. Stable | -3.19 | 0.008 | -5.53 | -0.85 |
| FACT-TOI change | | | | | | | | | | |
| | NIH consensus severity change | -0.35 | 0.630 | -1.79 | 1.09 | Improve vs. Stable | 0.52 | 0.688 | -2.01 | 3.05 |
| | | | | | | Worsen vs. Stable | -0.53 | 0.696 | -3.18 | 2.12 |
| | Clinician-assessed severity change | -1.61 | 0.040 | -3.15 | -0.07 | Improve vs. Stable | 2.43 | 0.052 | -0.02 | 4.89 |
| | | | | | | Worsen vs. Stable | -0.71 | 0.630 | -3.60 | 2.18 |
| | Patient-reported severity change | -4.11 | <0.001 | -5.44 | -2.78 | Improve vs. Stable | 4.85 | <0.001 | 2.54 | 7.16 |
| | | | | | | Worsen vs. Stable | -3.26 | 0.014 | -5.85 | -0.68 |
| FACT-G change | | | | | | | | | | |
| | NIH consensus severity change | -0.49 | 0.511 | -1.96 | 0.98 | Improve vs. Stable | 0.02 | 0.985 | -2.55 | 2.60 |
| | | | | | | Worsen vs. Stable | -1.34 | 0.331 | -4.05 | 1.37 |
| | Clinician-assessed severity change | -2.52 | 0.002 | -4.08 | -0.97 | Improve vs. Stable | 2.83 | 0.026 | 0.34 | 5.32 |
| | | | | | | Worsen vs. Stable | -2.30 | 0.122 | -5.23 | 0.62 |
| | Patient-reported severity change | -3.97 | <0.001 | -5.31 | -2.63 | Improve vs. Stable | 4.34 | <0.001 | 2.00 | 6.68 |
| | | | | | | Worsen vs. Stable | -3.46 | 0.009 | -6.06 | -0.86 |
| FACT-BMT change | | | | | | | | | | |
| | NIH consensus severity change | -0.56 | 0.558 | -2.43 | 1.31 | Improve vs. Stable | 0.64 | 0.702 | -2.65 | 3.92 |
| | | | | | | Worsen vs. Stable | -0.81 | 0.644 | -4.27 | 2.65 |
| | Clinician-assessed severity change | -2.98 | 0.004 | -4.97 | -0.99 | Improve vs. Stable | 3.92 | 0.016 | 0.74 | 7.10 |
| | | | | | | Worsen vs. Stable | -1.97 | 0.300 | -5.72 | 1.77 |
| | Patient-reported severity change | -5.21 | 0.000 | -6.93 | -3.50 | Improve vs. Stable | 5.92 | <0.001 | 2.93 | 8.90 |
| | | | | | | Worsen vs. Stable | -4.40 | 0.010 | -7.74 | -1.06 |

*All models adjusted for educational level and months since enrollment. QOL: quality of life; SF36: Short Form 36; PCS: physical component scale; MCS: mental component scale; FACT: Functional Assessment of Cancer Therapy; TOI: Trial Outcome Index; G, general score; BMT: Bone Marrow Transplant score.

severity. These results show that in approximately two of five visits, patients would report a clinically significant change in their QOL over each 6-month period, and that the only way to identify these patients would be to ask them directly.

One of the goals of this analysis was to identify the most parsimonious battery of QOL instruments or items that would capture the outcomes of interest. It appears that both the SF-36 and FACT-BMT perform well in capturing the patients' perspective, and fairly well in capturing physicians' assessment. Neither correlates well with changes in NIH global severity scores. The SF-36 has 36 items, the FACT-TOI has 24 items, the FACT-G has 27 items and the FACT-BMT has 37 items: an examination of Table 4 suggests that the battery could be condensed to collection of only the FACT-G without loss of discrimination capabilities.

Patients complain about the redundant nature of the surveys and would probably welcome completing only 27 instead of 73 items, reducing the anticipated time commitment from 12 minutes to 4 minutes.

The strengths of this study include the prospective assessment and standardized reporting of data on chronic GVHD activity according to the NIH Consensus criteria, clinicians' assessment, and patients' evaluation. One of the limitations is the amount of missing data. Our analysis of the patterns of missing data showed only that more educated participants were more diligent about performing study procedures; this finding is consistent with a large body of published data demonstrating a relationship between patients' adherence and education level.²⁵ Missing data tended to occur as the time since enrollment increased and after a longer interval between visits. We could not identify any

Table 5. Frequency of clinically meaningful change in QOL and weighted kappa statistics between changes in chronic GVHD severity and clinically meaningful changes in QOL.

| | Improvement | | | Stable | | | Worsening | | | Agreement | | |
|----------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|---------|
| | NIH | Clinician | Patient | NIH | Clinician | Patient | NIH | Clinician | Patient | NIH | Clinician | Patient |
| SF36-PCS | 93 (23%) | 90 (23%) | 85 (22%) | 228 (56%) | 224 (56%) | 221 (57%) | 84 (21%) | 83 (21%) | 82 (21%) | -0.04 | -0.05 | 0.24 |
| SF36-MCS | 92 (23%) | 92 (23%) | 86 (22%) | 234 (58%) | 228 (58%) | 225 (58%) | 79 (19%) | 77 (19%) | 77 (20%) | 0.04 | 0.12 | 0.18 |
| FACT-TOI | 86 (21%) | 84 (21%) | 80 (20%) | 256 (62%) | 253 (63%) | 250 (64%) | 68 (17%) | 65 (16%) | 63 (16%) | 0.01 | 0.13 | 0.22 |
| FACT-G | 89 (22%) | 88 (22%) | 80 (20%) | 242 (59%) | 237 (59%) | 237 (61%) | 76 (19%) | 74 (19%) | 73 (19%) | 0.04 | 0.18 | 0.21 |
| FACT-BMT | 107 (27%) | 105 (26%) | 100 (26%) | 204 (50%) | 201 (51%) | 200 (51%) | 95 (23%) | 92 (23%) | 89 (23%) | 0.04 | 0.15 | 0.26 |

QOL: quality of life; SF36: Short Form 36; PCS: physical component scale; MCS: mental component scale; FACT: Functional Assessment of Cancer Therapy; TOI: Trial Outcome Index; G: General score; BMT: Bone Marrow Transplant score.

other patient, transplant or chronic GVHD characteristics associated with missing data.

In summary, these data indicate that NIH and clinician-determined chronic GVHD severity assessments do not reflect the full QOL experience of people with chronic GVHD. Inclusion of patient-reported outcomes in the conduct of clinical trials and routine clinical practice provides a unique perspective that is not currently captured through any clinician-reported assessments. We recommend use of the FACT-G core survey rather than the SF-36 and the FACT-BMT subscale since the core survey would substantially decrease respondent burden without compromising

understanding of multi-dimensional QOL and sensitivity to change.

Authorship and Disclosures

The information provided by the authors about contributions from persons listed as authors and in acknowledgments is available with the full text of this paper at www.haematologica.org.

Financial and other disclosures provided by the authors using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are also available at www.haematologica.org.

References

- Duell T, van Lint MT, Ljungman P, Tichelli A, Socie G, Apperley JF, 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):184-92.
- Lee SJ, Flowers ME. Recognizing and managing chronic graft-versus-host disease. *Hematology Am Soc Hematol Educ Program.* 2008:134-41.
- Stewart BL, Storer B, Storek J, Deeg HJ, Storb R, Hansen JA, et al. Duration of immunosuppressive treatment for chronic graft-versus-host disease. *Blood.* 2004;104(12):3501-6.
- Socie G, Stone JV, Wingard JR, Weisdorf D, Henslee-Downey PJ, Bredeson C, et al. Long-term survival and late deaths after allogeneic bone marrow transplantation. Late Effects Working Committee of the International Bone Marrow Transplant Registry. *New Eng J Med.* 1999;341(1):14-21.
- 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):1621-8.
- Halyard MY, Ferrans CE. Quality-of-life assessment for routine oncology clinical practice. *J Support Oncol.* 2008;6(5):221-9, 233.
- Lee SJ, Fairclough D, Parsons SK, Soiffer RJ, Fisher DC, Schlossman RL, et al. Recovery after stem-cell transplantation for hematologic diseases. *J Clin Oncol.* 2001;19(1):242-52.
- 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, Petti AR, Van Lint MT, Lamparelli T, et al. Quality of life in 244 recipients of allogeneic bone marrow transplantation. *Br J Haematol.* 2000;110(3):614-9.
- Fraser CJ, Bhatia S, Ness K, Carter A, Francisco L, Arora M, 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):2867-73.
- 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):2334-43.
- Lee SJ, Kim HT, Ho VT, Cutler C, Alyea EP, Soiffer RJ, et al. Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant.* 2006;38(4):305-10.
- 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):319-27.
- Worel N, Biener D, Kalhs P, Mitterbauer M, Keil F, Schulenburg A, 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):619-26.
- Pidala J, Kurland B, Chai X, Majhail N, Weisdorf DJ, Pavletic S, et al. Patient-reported quality of life is associated with severity of chronic graft-versus-host disease as measured by NIH criteria: report on baseline data from the Chronic GVHD Consortium. *Blood.* 2011;117(17):4651-7.
- Filipovich AH, Weisdorf D, Pavletic S, Socie G, Wingard JR, Lee SJ, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. *Biol Blood Marrow Transplant.* 2005;11(12):945-56.
- McQuellon RP, Russell GB, Cella DF, Craven BL, Brady M, Bonomi A, 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):357-68.
- 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):247-63.
- Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston: The Health Institute, 1994.
- Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Boston: The Health Institute, 1993.
- Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34(3):220-33.
- Fleiss J, Cohen, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement.* 1973;33:613-9.
- Laird NM. Missing data in longitudinal studies. *Stat Med.* 1988;7(1-2):305-15.
- Administration UDOHaHSFaD. Guidance for Industry - Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2006 February.
- DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med Care.* 2004;42(3):200-9.