



Published in final edited form as:

Cancer. 2018 February 15; 124(4): 841–849. doi:10.1002/cncr.31089.

PROMIS Measures can be used to assess symptoms and function in long term Hematopoietic Cell Transplantation Survivors

Bronwen E Shaw, MD, PhD¹, Karen L. Syrjala, PhD², Lynn E Onstad², Eric J Chow, MD, MPH³, Mary E. Flowers, MD², Heather Jim, PhD⁴, K Scott Baker, MD MS², Sarah Buckley, MD⁵, Diane L Fairclough, DrPH, MDPH⁶, Mary M Horowitz, MD, MS¹, and Stephanie J. Lee, MD, MPH²

¹CIBMTR (Center for International Blood and Marrow Transplant Research), Department of Medicine, Medical College of Wisconsin, Milwaukee, WI

²Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

³Pediatric Hematology-Oncology, Fred Hutchinson Cancer Research Center, Seattle, WA

⁴Health Outcomes and Behavior, Moffitt Cancer Center., Tampa, FL

⁵Hematology/Oncology Fellowship Program, University of Washington, Seattle

⁶Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO

Abstract

Background—Patient-reported outcomes (PRO) in hematopoietic cell transplantation (HCT) survivors are well-characterized using established measures; however, there is little experience with the new, freely available Patient-Reported Outcomes Measurement Information System (PROMIS) measures in this population. The aim of this study was to report the performance of the PROMIS measures in the HCT setting compared with the commonly used Short Form 36 (SF36).

Methods—4,446 adult HCT survivors from Fred Hutchinson Cancer Research Center were mailed a survey that included: PROMIS Global Health (GH: 10 questions), PROMIS profile 29 (PROMIS-29) and SF36 as part of an annual follow-up survey.

Results—Both the SF36 and PROMIS measures were available for 1,634 (503 autologous, 1,131 allogeneic) HCT recipients. The overall response rate was 46%. The median time post-transplant for allogeneic and autologous recipients was 12.0 (range, 0.4-44.1) and 6.1 (range 0.4-30.1) years, respectively. Using the SF36 or PROMIS GH, overall physical functioning was somewhat lower than the general population but mental functioning was similar. Component and domain scores with similar content were strongly correlated by Pearson correlation coefficients: GH-physical and

Corresponding author: Bronwen E Shaw, CIBMTR, 9200 W. Wisconsin Avenue, Suite C5500, Milwaukee, WI, Phone: +1-414-805-0700, beshaw@mcw.edu, Fax: +1-414-805-0714.

Authorship: SJL designed the study. LO performed the statistical analysis. All authors contributed to the data interpretation, writing and review of the manuscript.

Disclosures: None of the authors have any relevant conflict of interest to report and all authors have approved the final version of the manuscript.

SF36 physical component summary for autologous ($r=0.82$) and allogeneic ($r=0.83$) recipients; and PROMIS-29 and SF36 physical function, pain and vitality/fatigue (allogeneic: 0.87, -0.82, -0.82 and autologous: 0.84, -0.82 and -0.81 recipients respectively). The correlation between the GH-mental and SF36 mental component summary score was lower (autologous: 0.70; allogeneic: 0.72).

Conclusions—Physical and mental symptoms and function in autologous and allogeneic HCT survivors can be adequately assessed by the PROMIS-29 and PROMIS-Global Health.

Keywords

Quality of Life; hematopoietic cell transplantation; SF36; PROMIS; patient-reported outcomes; convergent validity

Introduction

As the field of hematopoietic cell transplantation (HCT) matures and the number of long-term survivors grows¹, it is critical to acknowledge and understand the significant symptoms and functional deficits and late effects that are caused by the treatment itself. QOL is a dynamic, multidimensional construct that is assessed by patient self-report. Numerous patient-reported outcome (PRO) measures exist for the purpose of assessing symptoms and function. In fact, the multitude of available measures now represents a challenge to the field, with a variety of measures addressing the same concept or with one or more overlapping domains, widely varying quality, excessive length and complexity, and difficulty comparing findings across studies and conditions².

Both autologous (patient's own cells reinfused) and allogeneic (cells from a donor) HCT are intensive procedures. Failure of the HCT, with relapse of the original disease, is a devastating outcome for either HCT methodology, but many patients can be treated again with a second transplant or novel therapeutic agents. Due to the immunological disparity between the patient and donor cells, patients undergoing allogeneic HCT are also at risk of graft-versus-host-disease (GVHD), a multi-system disorder requiring high doses of immune suppressants for therapy and well recognized to be associated with significant deficits in symptoms and function. For patients who do not relapse or develop GVHD, other late-effects develop at a rate higher than the general population, including endocrine (e.g. diabetes, hypothyroidism), cardiac (e.g. metabolic syndrome) and subsequent malignancies³. These conditions can all result in deficits in symptoms and function post-HCT, which for some patients, may be long lasting, however a large proportion of patients will achieve their baseline symptoms and function at 1-2 years post-HCT⁴. Factors which have previously been associated with worse symptoms and function include age, gender and certain characteristics of the transplant procedures (e.g. chemo-radiotherapy, donor type) which increase the risk of GVHD^{5,6}. Both negative and positive impacts on a patient's psychological wellbeing have been described in the post-HCT setting⁷.

Our group recently published a review and opinion piece⁸ calling for harmonization in the use of PRO measures in the setting of HCT. That article provided several general recommendations, including the desirability of developing a standard core set of domains of

interest and of using a free and easily accessible system that is versatile and has a low respondent burden.

The Patient-Reported Outcomes Measurement Information System (PROMIS®), is a set of person-centred measures housed under the healthmeasures.net platform. Developed in 2004, PROMIS used modern psychometric theory to standardize PRO assessment for use in both clinical research and health care delivery settings. Extensive psychometric testing has since been done in large community and clinical sample (including cancer patients) to validate PROMIS item banks⁹⁻¹¹ A key feature of PROMIS is that all scores are generic (not disease-specific) and thus applicable to use in the setting of HCT where multiple different diseases may be represented in the patient population. PROMIS offers numerous benefits to address current symptom and functional measurement challenges, including that it is free to use, easily accessible (via multiple electronic and paper-based platforms), and flexible. All PROMIS scores use a standardized t-score metric against normative data for the U.S. general population. All PROMIS measures within a domain can be scored and compared on the same t-score metric, regardless of number and items administered. This allows for comparisons across studies and disease-specific populations that may use different PROMIS measures. Legacy measures (including the SF-36) have been linked to specific PROMIS domains using the PROsetta Stone project allowing all measures to be used and interpreted within the same t-score scale (<http://www.prosetastone.org/Pages/default.aspx>). Additionally, most PROMIS measures can be delivered as static short forms or through computer adaptive testing (CAT)¹², an innovation that can decrease responder burden by measuring symptoms and function using fewer questions, because it tailors subsequent questions based on a patient's previous answers. PROMIS CAT has been developed using data elements that will allow for ease of integration with the electronic medical record². Finally, PROMIS measures are validated in multiple languages (<http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/available-translations>).

Despite potential logistic and scientific benefits, there is limited experience with the PROMIS measures in HCT¹³⁻¹⁵ and only a single study has investigated correlations between PROMIS and other PRO measures in this population¹⁴.

Therefore, the aim of this study was to determine whether the PROMIS measures perform well in the assessment of general symptoms and function in the post-HCT setting by investigating the correlation between the PROMIS and the commonly utilized SF36 measures.

Recipients, materials and methods

Study design

This is a cross-sectional study of symptoms and function in HCT survivors.

Study population

All surviving adult patients transplanted at Fred Hutchinson Cancer Research Center (FHCRC) in Seattle, Washington, USA are mailed annual surveys, regardless of their disease

status. HCT recipients receive these surveys on approximately their transplant anniversaries and are sent one reminder if the survey is not returned within one month. Patients provide written informed consent at the time of their transplants, and only 3.7% of autologous recipients and 2.1% of allogeneic recipients decline, later stop participating, or are not allowed to be contacted (e.g., due to incarceration). An online option that mirrors the paper version is available to all survey recipients. The annual survey consists of a core set of questions along with a modifiable research module that is distributed for one year. The July 2015-June 2016 module included PROMIS measures and was mailed to a total of 4446 adults, including 93 whose family subsequently reported that the patient had died since last contact.

Measures

The SF36 is a 36-item scale constructed to survey health status and quality of life¹⁶. It consists of eight domains, which are the weighted sums of relevant questions. The items use Likert-type scales, some with 5 or 6 points and others with 2 or 3 points. Each scale of the eight domains is transformed into a standardized 0-100 z-score (with a mean of 50 and standard deviation of 10). Lower scores indicate worse symptoms and function. The eight domains are: vitality (VT), physical functioning (PF), bodily pain (BP), general health perceptions (SF36 GH), physical role functioning (RP), emotional role functioning (RE), social role functioning (SF), and mental health (MH). Two composite scores can also be generated: the physical component summary (PCS) and the mental component summary (MCS).

PROMIS consists of item banks with a variable number of questions that can be combined to form multi-item measures of varying length and complexity⁹. The items use Likert-type scales with 5 points. Because each item is mapped onto a common metric, measures of differing length can be compared to each other. Similar to the SF36, PROMIS uses a T-score standardized metric in which 50 is the mean of a relevant reference population and 10 is the standard deviation (SD) of that population. For negatively-worded concepts like fatigue, a higher T-score represents greater fatigue, while for a positively-worded concept like physical function, a higher T-score reflects higher (better) physical function. We used two PROMIS instruments. The PROMIS Global Health Scale (Global-10) is an instrument that consists of 10 items, and can be scored into a Global Physical Health component (GH-physical) and Global Mental Health component (GH-mental). Every question in a domain must be answered for that domain to be scored. The PROMIS GH is conceptually similar to the SF-12¹⁷. The PROMIS-29 profile contains 7 PROMIS domains with 4 questions each (short-forms), and one pain intensity question. PROMIS short forms in the PROMIS-29 include: Depression (4a), Anxiety (4a), Physical Function (PF4a), Pain Interference (4a), Fatigue (4a), Sleep Disturbance (4a), and Ability to Participate in Social Roles and Activities v2 (4a).

Clinical data

Clinical and socio-demographic variables were abstracted from the patient's chart in the institutional database. These included: age, gender, year of transplantation, conditioning intensity, donor type, cell source and disease relapse. Patient self-report was used for post-

transplant events including: graft-versus-host-disease and co-morbidity burden (pulmonary disease, avascular necrosis, adrenal insufficiency and diabetes).

Ethical permission

The study has ethical approval from the FHCRC Institutional Review Board and all survey respondents signed informed consent.

Data sharing

Our results provide benchmarks for PROMIS-29 and Global Health scores reported by long-term HCT survivors, which can be referenced by future studies; scores for subsets are available by contacting the authors.

Statistical Methods

Descriptive statistics for the population were calculated and compared between respondents and non-respondents using the chi-square test and t-test for categorical and continuous variables, respectively. Subsequent analyses were restricted to those who returned both a scorable SF36 (>50% of questions answered) and the PROMIS GH measure (as a minimum). Since some patients did not answer every question in the Global Health instrument, there are very slight differences in the analytic populations. Cronbach's alpha¹⁸ was calculated for each PROMIS measure to evaluate internal consistency. Differences between the SF36 PCS and GH-physical values and between the SF36 MCS and GH-Mental values were compared using paired t-tests. A p-value <0.01 was considered statistically different and a difference of 5 points (half a standard deviation) was considered clinically meaningful. In addition, relationships between global scores and domains across the SF36 and PROMIS measures were examined using Pearson's correlation coefficients. Cohen's criterion was used to interpret the magnitude of correlation coefficients ($r < 0.3 = \text{small}$; $r > 0.3$ and $< 0.5 = \text{medium}$; and $r > 0.5 = \text{large}$)¹⁹. Additionally, PROsetta Stone[®] linkages were used to calculate PROMIS-29 measure scores from SF36 raw scores (for physical function, pain and vitality/fatigue), and extrapolated PROMIS-29 scores were then compared to the actual PROMIS-29 domain t-scores using Pearson's correlation coefficients.

Multivariable analyses were performed to investigate the relationships between the composite scores for PROMIS GH and SF36 (physical and mental) and transplant related factors (age, gender, year of transplant, conditioning intensity, comorbidity burden and relapse) in all recipients. In addition, donor type, cell source and chronic graft versus host disease (cGVHD) were considered in allogeneic recipients.

Results

Both the SF36 and PROMIS measures were available for 1,634 (503 autologous, 1,131 allogeneic) HCT recipients, and an additional 382 (119 autologous, 263 allogeneic) had only one set of measures, for an overall response rate of 46%. The median times since transplant, at the time of the survey, for allogeneic and autologous recipients were 12.0 (range, 0.4-44.1) and 6.1 (range 0.4-30.1) years, respectively. The larger numbers and longer survival of the allogeneic recipients reflects the early emphasis on allogeneic transplantation

at FHCRC. Table 1a and b show the pre-transplant characteristics for this study population as well as the differences between respondents and non-respondents. Allogeneic non-respondents were younger, were slightly longer out from transplant, and more likely to have received myeloablative conditioning, TBI-based regimens and bone marrow grafts than respondents. Autologous non-respondents were younger and more likely to receive TBI-based conditioning than respondents.

Patient-reported outcomes scores

Figure 1 shows the component score distributions for the SF36 (PCS and MCS) and the PROMIS GH (physical and mental) in allogeneic and autologous HCT recipients. In allogeneic patients (n=1,131), the median SF36 PCS was 49.2 (IQR 39.2-55.7) and the median MCS was 54.3 (IQR 46.2-58.5). The GH-physical had a median of 50.8 (IQR 42.3-57.7) and the GH-mental had a median of 50.8 (IQR 43.5-56.0). In autologous patients (n=503), the median SF36 PCS was 46.9 (IQR 37.5-54.4) and the median MCS was 54.6 (IQR 46.7-58.9). The GH-physical had a median of 47.7 (IQR 42.3-54.1) and the GH-mental had a median of 50.8 (IQR 45.8-56.0). The mean difference between the SF36 PCS and GH-physical was -3.5 (95% CI -3.9, -3.1, p<0.001) and -3.9 (95% CI -4.4, -3.3, p<0.001) in allogeneic and autologous recipients, respectively. Thus, the SF36 PCS yields a statistically but not clinically meaningful lower physical score than the PROMIS GH-physical measure. The mean difference between the SF36 MCS and GH-mental in allogeneic and autologous recipients was 0.6 (95% CI 0.1, 1.0, p=0.008) and 1.0 (95% CI 0.3, 1.6, p=0.004), respectively which again was statistically different but not clinically meaningful. Figures 2a and b show the domain score distributions for the SF36 and the PROMIS29 in allogeneic and autologous HCT recipients, respectively.

Correlations between SF36 and PROMIS Global-10 measures

Table 2a shows the correlations between the PROMIS GH and the SF36 PCS and MCS component scores. The correlations between the GH-physical and PCS were strong (Pearson correlation coefficient in allogeneic: 0.83 and autologous: 0.82). The correlations were slightly lower but still strong for the mental measures (Pearson correlation coefficient in allogeneic: 0.72, autologous: 0.70). Interestingly, the SF36 physical and mental component scores were not highly correlated (0.13-0.23) whereas the GH physical and mental components were more closely correlated (0.69-0.72). The PROMIS GH domains had high internal reliability, ranging from α =0.83-0.97.

PROMIS-29 and SF36 Comparisons

Three domains appear to measure similar constructs within the SF36 and PROMIS29: physical function, pain and vitality/fatigue. Strong correlations were seen between similar constructs (Pearson correlation coefficient in allogeneic (0.87, -0.82, -0.82, respectively) and autologous (0.84, -0.82, -0.81, respectively) HCT recipients (Table 2b). Using the functionality provided in PROsetta stone, we found statistically significant correlations between the domain scores mapped from the SF36 measures and scores from the PROMIS measures, as well as the SF36 MH and PROMIS Anxiety and Depression (data not shown).

Association of clinical factors with symptoms and function

Results of the multivariable analyses are shown in Table 3a and b. In allogeneic recipients, older age at the time of the survey and less time since transplant were associated with worse physical scores. Younger patients reported worse mental health measured by the SF36. When considering post-transplant factors, a history of chronic GVHD or of self-reported comorbidities were associated with significantly worse physical and mental functioning, whether measured by either the PROMIS GH or the SF36 (although not statistically significant as measured by the SF36 MCS for comorbidities). A history of relapse was not significantly associated with PRO in allogeneic recipients. In autologous recipients, younger patients reported worse mental health (statistically significant on the SF36 only). Older age, shorter time post-transplant, presence of comorbidities and relapse were all associated with statistically significantly lower scores on the SF36 PCS. Although the associations with these factors were all directionally consistent when measured by the GH, only comorbidities were statistically significant.

Discussion

These data provide information about symptoms and function for a very large number of long-term HCT survivors who received allogeneic or autologous transplantation in a single transplant center. Scores show that the overall symptoms and function, in those patients who do respond to the survey, is similar to the general population. They also show that PROs as measured by PROMIS are strongly correlated with those from a legacy measure, the SF36, either by scoring the patient responses on the SF36 or by linking through PROsetta stone. Internal reliability of the PROMIS measures was high. These findings support the proposal that the PROMIS measures are reliable and can be used as an alternative PRO measure in the setting of HCT⁸.

Among respondents, physical health was found to be approximately half a standard deviation below the general population norm (slightly lower for autologous recipients than allogeneic recipients). This is consistent with the HCT literature where continued long-term (mild to moderate) impairment in physical function, relative to comparison groups, is reported, even at 5-10 years post-HCT²⁰. We found that post-HCT events had a significant association with symptoms and function. The presence of self-reported chronic conditions (presence of at least one of pulmonary disease, avascular necrosis, adrenal insufficiency and diabetes or chronic GVHD)^{5,21} was highly correlated with physical function, as has been previously reported²⁰. The presence of relapse was associated with worse physical function in autologous, but not allogeneic patients. Given the long time post-transplant there may be fewer allogeneic patients experiencing issues related to relapse at the time of self-report (compared to autologous patients)⁵, while patients who had relapsed earlier may be less likely to remain alive. Alternatively, ongoing treatment or maintenance therapy for relapse may affect symptoms and function.

While there does appear to be a general tendency for the SF-36 PCS to produce scores that are lower than the GH-physical by 3-4 points, these composite scores were strongly correlated. Our ability to screen for impaired physical function using the short PROMIS GH 10-item questionnaire would be beneficial in terms of reducing respondent burden and thus,

potentially, increasing retention and participation. Respondent burden remains a significant concern and potential barrier in PRO collection. A recent meta-analysis of 20 studies examining the impact of respondent burden found a significant association between a lower response rate and longer questionnaire length ($P = 0.0001$)²². Wood et al¹⁵ reported, in an early post-HCT cohort, that the median time to complete the PROMIS GH measure was 3 minutes (using an electronic PRO system) further supporting this benefit.

Only two previous studies investigated the PROMIS measures in HCT patients. Wood et al¹⁵ administered the PROMIS GH to 32 HCT recipients at baseline and then weekly until day 100 post-HCT. Although the numbers were too low to draw robust conclusions, they found that the PRO scores varied predictably over time, with changes in the physical scores mirroring symptoms. In addition, they reported correlations between post-HCT PROMIS GH scores and pedometry data and found associations between fewer average daily steps and worse scores in all physical and social domains¹³. They did not include any other measures as comparators. A cross-sectional study in 136 HCT recipients¹⁴ provided evidence that the PROMIS Cancer Fatigue Short Form was reliable and valid in this population compared to legacy measures including the vitality subscale of the SF36 and the Fatigue Symptom Inventory (FSI). Scores on the PROMIS Cancer Fatigue Short Form were positively associated with fatigue severity, and fatigue disruptiveness measured by FSI ($p < 0.0001$), and lower PROMIS scores were associated with higher vitality scores measured by SF36 ($p < 0.0001$).

The pattern for mental scores differed from the patterns in the physical domains in our study. The mean mental scores were near the population norm for all measures, with a tendency for the PROMIS GH score to be slightly, but statistically significantly, lower than the SF36. Multivariate analysis showed fewer clinical and demographic factors were significantly associated with the mental scores, although, as with physical function, post-HCT events had significant impact on mental scores in the allogeneic setting. In addition, younger patients reported worse mental functioning. These findings suggest that post-HCT survivors, particularly younger patients and those with ongoing clinical issues, should be regularly assessed for anxiety and depression, with a low threshold for offering interventions. The general consistency in the direction of effect of clinical predictors provides some indirect support about the comparability of the SF36 and PROMIS GH.

Our study has several limitations. Fewer than half of patients completed the PRO and a small proportion returned only one or other of the measures, precluding inclusion in the analysis. It is unlikely that response bias, resulting from these missing data, would affect conclusions about the correlation between the PROMIS and SF36 measures, however, it is possible that the associations found in the multivariate analysis may not be generalizable based on demographic differences between those who did and did not return the PROs. Future interventions to encourage participation, particularly from underrepresented populations (e.g. younger patients), are needed. Secondly, because it is cross-sectional, as in previous HCT publications¹⁴, the test-retest reliability and sensitivity to change of the PROMIS measures could not be assessed. Future longitudinal studies are necessary to confirm that the PROMIS measures perform as expected in the setting of long-term HCT survivorship. Third, the measures were delivered as static forms, and time studies to address respondent burden

were not done specifically for the PROMIS forms. Future studies using CAT should be performed to understand the degree to which this approach decreases respondent burden since there could be trade-offs in terms of ease of survey administration for staff. Finally, although the PROMIS measures have been extensively validated in general, content validity and internal structure have not been specifically addressed in this population, and further research in this area is recommended.

In conclusion, we found that the PROMIS measures performed as expected when used alongside a well-understood legacy measure, the SF36, in a cross-sectional population of HCT recipients. In this setting, we believe that the short GH measure provides an attractive screening tool with the benefits in delivery, logistics and cost that the PROMIS measures offer, to reduce respondent burden and enhance patient participation and retention.

Acknowledgments

Funding was provided by NCI CA18029.

The authors thank Roxanne Jensen Ph.D. for helpful input.

References

1. Majhail NS, Tao L, Bredeson C, et al. Prevalence of hematopoietic cell transplant survivors in the United States. *Biol Blood Marrow Transplant*. 2013; 19:1498–501. [PubMed: 23906634]
2. Bevens M, Ross A, Cella D. Patient-Reported Outcomes Measurement Information System (PROMIS): efficient, standardized tools to measure self-reported health and quality of life. *Nurs Outlook*. 2014; 62:339–45. [PubMed: 25015409]
3. 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; 117:4651–7. [PubMed: 21355084]
4. Mosher CE, Redd WH, Rini CM, et al. Physical, psychological, and social sequelae following hematopoietic stem cell transplantation: a review of the literature. *Psychooncology*. 2009; 18:113–27. [PubMed: 18677717]
5. Pidala J, Anasetti C, Jim H. Quality of life after allogeneic hematopoietic cell transplantation. *Blood*. 2009; 114:7–19. [PubMed: 19336756]
6. Wingard JR, Huang IC, Sobocinski KA, et al. Factors associated with self-reported physical and mental health after hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2010; 16:1682–92. [PubMed: 20685400]
7. Andrykowski MA, Bishop MM, Hahn EA, et al. Long-term health-related quality of life, growth, and spiritual well-being after hematopoietic stem-cell transplantation. *J Clin Oncol*. 2005; 23:599–608. [PubMed: 15659507]
8. Shaw BE, Lee SJ, Horowitz MM, et al. Can we agree on patient-reported outcome measures for assessing hematopoietic cell transplantation patients? A study from the CIBMTR and BMT CTN. *Bone Marrow Transplant*. 2016; 51:1173–9. [PubMed: 27159181]
9. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010; 63:1179–94. [PubMed: 20685078]
10. Liu H, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol*. 2010; 63:1169–78. [PubMed: 20688473]
11. Rothrock NE, Hays RD, Spritzer K, et al. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported

- Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2010; 63:1195–204. [PubMed: 20688471]
12. Cella D, Gershon R, Lai JS, et al. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res*. 2007; 16(Suppl 1):133–41. [PubMed: 17401637]
 13. Bennett AV, Reeve BB, Basch EM, et al. Evaluation of pedometry as a patient-centered outcome in patients undergoing hematopoietic cell transplant (HCT): a comparison of pedometry and patient reports of symptoms, health, and quality of life. *Qual Life Res*. 2016; 25:535–46. [PubMed: 26577763]
 14. Cessna JM, Jim HS, Sutton SK, et al. Evaluation of the psychometric properties of the PROMIS Cancer Fatigue Short Form with cancer patients. *J Psychosom Res*. 2016; 81:9–13. [PubMed: 26800633]
 15. Wood WA, Deal AM, Abernethy A, et al. Feasibility of frequent patient-reported outcome surveillance in patients undergoing hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2013; 19:450–9. [PubMed: 23253558]
 16. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992; 30:473–83. [PubMed: 1593914]
 17. 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; 34:220–33. [PubMed: 8628042]
 18. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011; 2:53–55. [PubMed: 28029643]
 19. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
 20. Pidala J, Anasetti C, Jim H. Health-related quality of life following haematopoietic cell transplantation: patient education, evaluation and intervention. *Br J Haematol*. 2010; 148:373–85. [PubMed: 19919651]
 21. Lee SJ, Kim HT, Ho VT, et al. Quality of life associated with acute and chronic graft-versus-host disease. *Bone Marrow Transplant*. 2006; 38:305–10. [PubMed: 16819438]
 22. Rolstad S, Adler J, Ryden A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value Health*. 2011; 14:1101–8. [PubMed: 22152180]

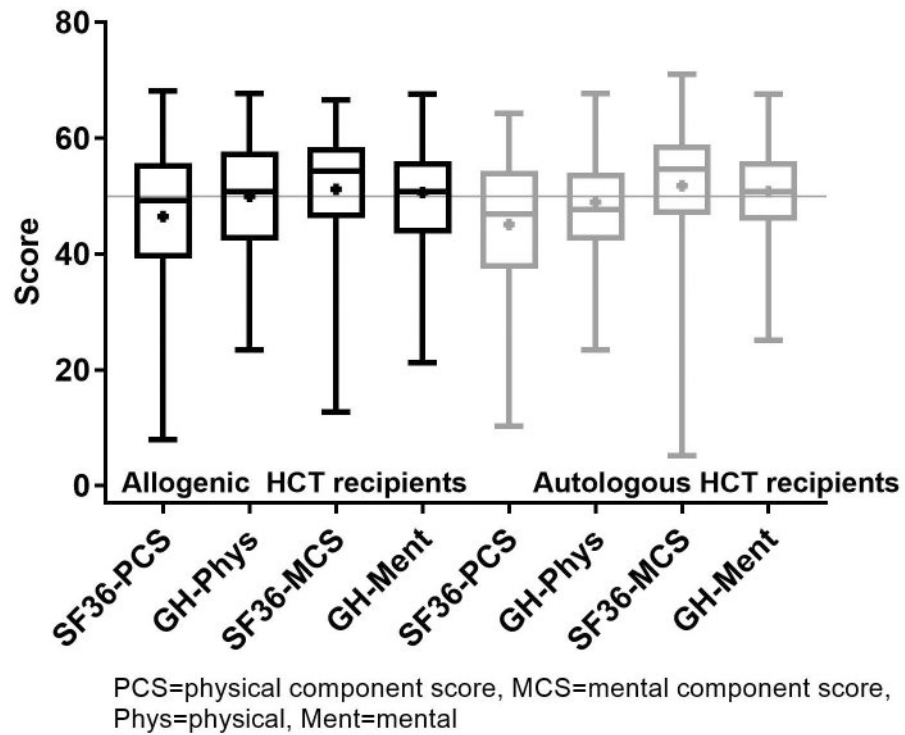
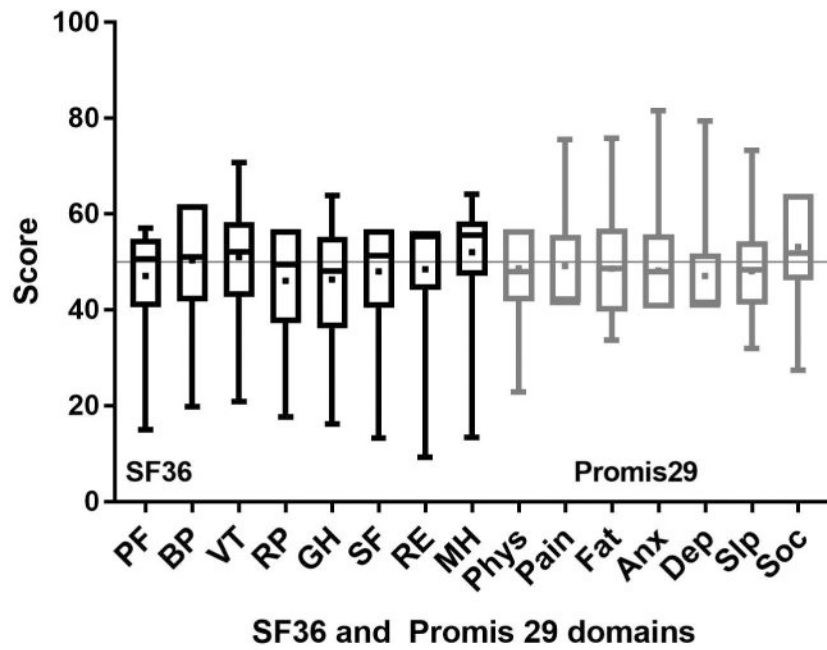
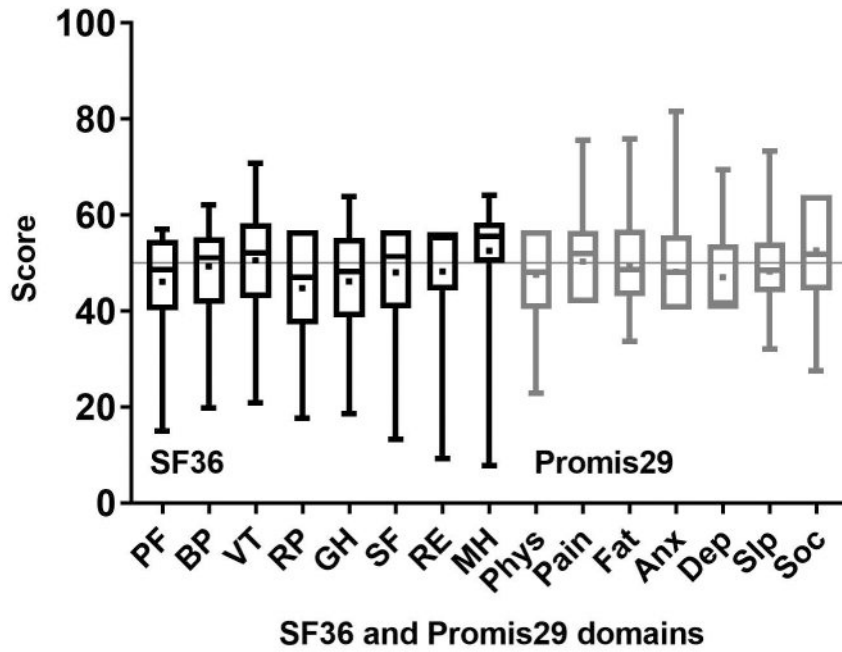


Figure 1. SF36 and PROMIS component score distribution. Data are displayed as mean (square), median (line inside the box), Interquartile range (IQR) (top & bottom edges), and extreme values (whiskers).



Vitality (VT), physical functioning (PF), bodily pain (BP), general health perceptions (SF36 GH), physical role functioning (RP), emotional role functioning (RE), social role functioning (SF), mental health (MH), physical function (Phys), Fatigue (Fat), Anxiety (Anx), Depression (Dep), Sleep (Slp), Social roles (Soc)



Vitality (VT), physical functioning (PF), bodily pain (BP), general health perceptions (SF36 GH), physical role functioning (RP), emotional role functioning (RE), social role functioning (SF), mental health (MH), physical function (Phys), Fatigue (Fat), Anxiety (Anx), Depression (Dep), Sleep (Slp), Social roles (Soc)

Figure 2a. SF36 and PROMIS29 domain score distribution in allogeneic HCT recipients. Data are displayed as mean (square), median (line inside the box), Interquartile range (IQR) (top & bottom edges), and extreme values (whiskers). SF36: lower scores indicate worse symptoms and function. PROMIS: for negatively-worded concepts a higher T-score represents more of the concept being measured (depression, anxiety, pain interference, fatigue, sleep disturbance), for positively-worded concepts a higher T-score reflects more (better) of the concept being measured (physical function, social interaction).
 SF36 and PROMIS29 domain score distribution in autologous HCT recipients. Data are displayed as mean (square), median (line inside the box), IQR (top & bottom edges), and extreme values (whiskers). SF36: lower scores indicate worse symptoms and function. PROMIS: for negatively-worded concepts a higher T-score represents more of the concept being measured (depression, anxiety, pain interference, fatigue, sleep disturbance), for positively-worded concepts a higher T-score reflects more (better) of the concept being measured (physical function, social interaction).

Table 1
a: Relevant pre-transplant characteristics in allogeneic HCT recipients by respondents and non-respondents

	Responded			P-value
	Yes (n=1131)	Partial* (n=263)	No (n=1633)	
Sex, n (%)	532 (47.0)	117 (44.5)	707 (43.3)	0.15
Female				
Age	56.2 (12.9)	57.1 (14.5)	48.4 (14.1)	<0.001
Mean years (SD)				
Donor relationship, n (%)	519 (45.9)	137 (52.1)	791 (48.4)	0.02
Matched related				
Mismatched related	53 (4.7)	10 (3.8)	105 (6.4)	
Haplo-identical related	30 (2.7)	4 (1.5)	31 (1.9)	
Matched unrelated	376 (33.2)	83 (31.6)	462 (28.3)	
Mismatched unrelated	108 (9.5)	15 (5.7)	147 (9.0)	
Cord	31 (2.7)	10 (3.8)	70 (4.3)	
Syngeneic	12 (1.1)	4 (1.5)	21 (1.3)	
Bone marrow	521 (46.1)	113 (43.0)	917 (56.2)	<0.001
Peripheral blood	579 (51.2)	140 (53.2)	645 (39.5)	
Yes	894 (79.0)	197 (74.9)	1408 (86.2)	<0.001
Myeloablative, n (%)				
Yes	383 (33.9)	94 (35.7)	724 (44.3)	<0.001
High dose TBI, n (%)				
Yes	13.7 (10.2)	14.1 (10.4)	15.9 (10.6)	<0.001
Years since HCT				
Mean years (SD)				

* Patients who returned the SF36 or the PROMIS GH, but not both

	Responded			P-value
	Yes (n=503)	Partial* (N=119)	No (N=797)	
Sex, n (%)	234 (46.5)	45 (37.8)	337 (42.3)	0.14
Female				
Age	61.1 (12.2)	63.6 (9.3)	55.3 (13.7)	<0.001
Mean years (SD)				
Peripheral blood	465 (92.4)	114 (95.8)	745 (93.5)	0.40
Graft source, n (%)				
Yes	503 (100)	119 (100)	797 (100)	1.0
Myeloablative, n (%)				
Yes	86 (17.1)	14 (11.8)	178 (22.3)	0.006
High dose TBI, n (%)				
Mean years (SD)	8.5 (7.5)	7.6 (6.8)	8.4 (7.2)	0.45
Year since HCT				

* Patients who returned the SF36 or the PROMIS GH, but not both

Table 2
a: Correlations between the components of the SF36 and the PROMIS GH (Pearson correlation coefficients)

	SF36_MCS		GHPhys_T		GHMent_T	
	Allo	Auto	Allo	Auto	Allo	Auto
SF36_PCS SF36 Physical (0-100)	0.23***	0.13*	0.83***	0.82***	0.52***	0.47***
SF36_MCS SF36 Mental (0-100)			0.49***	0.45***	0.72***	0.70***
GHPhys_T PROMIS-GH Physical: T score					0.73***	0.69***
*p<0.01						
**p<0.001						
***p<0.0001						

b: Correlations between the SF36 and PROMIS29 domains (Pearson correlation coefficients)

	PROMIS-29 Phys		PROMIS-29 Pain		PROMIS-29 Fatigue	
	Allo	Auto	Allo	Auto	Allo	Auto
SF36_PF SF36 Physical Function (0-100)	0.87***	0.84***				
SF36_BP SF36 Bodily pain (0-100)			-0.82***	-0.82***		
SF36_Vitality SF36 Vitality (0-100)					-0.82***	-0.81***

* p<0.01

** p<0.001

*** p<0.0001

Table 3
a: PRO multivariate analysis results among allogeneic transplant recipients

	SF36-PCS		GH-physical		SF36-MCS		GH-mental	
	β^a	p-value	β	p-value	β	p-value	β	p-value
Current age	-0.15	<0.0001	-0.05	0.03	0.08	0.003	0.01	0.60
Female	0.05	0.95	-0.003	1.0	0.53	0.43	-0.007	0.99
Years since transplant	0.23	<0.0001	0.13	0.008	0.04	0.45	0.10	0.06
Co-morbidity ^b	-5.77	<0.0001	-4.53	<0.0001	-1.24	0.11	-2.82	<0.0001
Relapse	-1.38	0.28	-1.25	0.26	0.64	0.60	0.30	0.79
BM (vs PBSC)	1.10	0.33	0.77	0.43	0.98	0.36	0.74	0.47
MA vs RIC	0.73	0.46	0.88	0.30	-0.99	0.30	-0.61	0.49
Chronic GVHD	-2.73	0.0006	-2.78	<0.0001	-2.51	0.0009	-2.64	0.0002
Donor type		0.35		0.39		0.22		0.60
Cord blood	0.06		2.25		-1.61		1.55	
Related match	-0.98		-0.23		-0.68		0.50	
Related mismatch	0.16		-1.06		-3.41		0.61	
Unrelated match	0.76		1.07		0.22		1.29	
Unrelated mismatch	0.15		-0.16		-1.80		-0.36	

BM=bone marrow, PBSC=peripheral blood stem cells, MA=myeloablative, RIC=reduced intensity conditioning, GVHD=graft versus host disease

^a β =regression coefficient

^bCo-morbidity= Presence of at least one of pulmonary disease, avascular necrosis, adrenal insufficiency and diabetes

b: PRO multivariate analysis results among autologous transplant recipients

	SF36-PCS		GH-physical		SF36-MCS		GH-mental	
	β^a	p-value	β	p-value	β	p-value	β	p-value
Current age	-0.10	0.01	-0.02	0.46	0.18	<0.0001	0.05	0.11
Female	-0.79	0.41	-1.16	0.15	-1.96	0.03	-1.98	0.01
Years since transplant	0.24	0.0002	0.09	0.10	-0.02	0.80	0.03	0.59
Co-morbidity ^b	-4.20	0.0006	-3.93	0.0001	-0.66	0.56	-1.98	0.051
Relapse	-5.68	0.004	-2.67	0.10	0.88	0.62	-0.62	0.70

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

β_g = regression coefficient

Co-morbidity = Presence of at least one of pulmonary disease, avascular necrosis, adrenal insufficiency and diabetes