



Published in final edited form as:

Biol Blood Marrow Transplant. 2014 March ; 20(3): 337–344. doi:10.1016/j.bbmt.2013.11.025.

Pulmonary symptoms measured by the NIH Lung score predict overall survival, non-relapse mortality and patient-reported outcomes in chronic GVHD

Jeanne Palmer¹, Kirsten Williams², Yoshihiro Inamoto³, Xiaoyu Chai³, Paul J. Martin³, Linus Santo Tomas⁴, Corey Cutler⁵, Daniel Weisdorf⁶, Brenda F. Kurland⁷, Paul A. Carpenter³, Joseph Pidala⁸, Steven Z. Pavletic⁹, William Wood¹⁰, David Jacobsohn², Sally Arai¹¹, Mukta Arora⁶, Madan Jagasia¹², Georgia B. Vogelsang¹³, and Stephanie J. Lee³

¹Division of Hematology/Oncology, Medical College of Wisconsin, Milwaukee, WI

²Division of Blood and Marrow Transplantation, Children's National Medical Center, Center for Cancer and Blood Disorders, Washington, DC

³Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

⁴Division of Pulmonary Medicine, Medical College of Wisconsin, Milwaukee, WI

⁵Hematologic Malignancies, Dana-Farber Cancer Institute, Boston, MA

⁶Blood and Marrow Transplant Program, University of Minnesota, Minneapolis, MN

⁷Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA

⁸Blood and Marrow Transplantation, Moffitt Cancer Center, Tampa, FL

⁹National Cancer Institute, National Institutes of Health, Bethesda, MD

¹⁰Linenberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

¹¹Division of Blood and Marrow Transplantation, Stanford University Medical Center, Stanford, CA

¹²Hematology and Stem Cell Transplant Program, Vanderbilt University Medical Center, Nashville, TN

¹³Oncology, Johns Hopkins Hospital, Baltimore, MD

Abstract

© 2013 The American Society for Blood and Marrow Transplantation. Published by Elsevier Inc. All rights reserved.

Correspondence: Stephanie J Lee, M.D., M.P.H., Fred Hutchinson Cancer Research Center, D5-290, PO Box 19024, Seattle, WA 98109-1024; Telephone: (206)-667-6190; Fax: (206)-667-1034; sjlee@fhcrc.org.

Authorship

Contribution: JP, XC, SJL designed the study, collected and analyzed data, and wrote the paper. XC and BFK performed the statistical analysis and wrote the paper; KW, YI, PJM, LS, CC, DW, PAC, JP, SZP, WW, DJ, SA, MA, MJ, and GBV collected data and wrote the paper. All authors critically revised the manuscript for important intellectual content and approved the manuscript to be published.

Conflict-of-interest disclosure

The authors declare no competing financial interests related to this study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The 2005 NIH Consensus Conference recommended assessment of lung function in patients with chronic graft-versus-host disease (GVHD) by both pulmonary function tests (PFTs) and assessment of pulmonary symptoms. We tested whether pulmonary measures were associated with non-relapse mortality (NRM), overall survival (OS) and patient reported outcomes (PRO). Clinician and patient-reported data were collected serially in a prospective, multicenter observational study. Available PFT data were abstracted. Cox regression models were fit for outcomes using a time-varying covariate model for lung function measures and adjusting for patient and transplant characteristics and non-lung chronic GVHD severity. A total of 1591 visits (496 patients) were used in this analysis. The NIH symptom-based lung score was associated with NRM ($p=0.02$), overall survival ($p=0.02$), patient-reported symptoms ($p<0.001$) and functional status ($p<0.001$). Worsening of NIH symptom-based lung score over time was associated with higher NRM and lower survival. All other measures were not associated with OS or NRM, although some were associated with patient-reported lung symptoms. In conclusion, the NIH symptom-based lung score of 0–3 is associated with NRM, OS, and PRO measures in patients with chronic GVHD. Worsening of the NIH symptom-based lung score was associated with increased mortality.

Introduction

Pulmonary dysfunction causes significant morbidity and mortality after allogeneic hematopoietic cell transplantation (HCT). Symptoms may include shortness of breath with exertion, cough, or wheezing. Routine screening with pulmonary function tests (PFTs) can detect lung function abnormalities before they become symptomatic. Pulmonary dysfunction is characterized as obstructive when the FEV1 is less than 80% of expected and FEV1/FVC <0.70 . Restrictive lung disease is based on decrease in total lung capacity, and is suggested when the FEV1 or FVC is less than 80% expected and the FEV1/FVC ratio is >0.70 . Some patients have dysfunction of oxygen/carbon dioxide exchange as measured by a decrease in the diffusing capacity of carbon monoxide (DLCO). Multiple studies have shown that both symptomatic and asymptomatic pulmonary complications that occur later in the transplant course are frequently associated with graft-versus-host disease (GVHD).^{1–8} Bronchiolitis obliterans syndrome (BOS) is the best-defined pulmonary manifestation of chronic GVHD.⁹ Bronchiolitis obliterans syndrome is diagnosed in approximately 6% of all HCT recipients, and in approximately 16% of patients with chronic GVHD.¹⁰ Factors reported to predict BOS include chronic GVHD,^{2, 4, 10–16} use of methotrexate as GVHD prophylaxis,¹² the use of busulfan as part of the conditioning regimen,^{3, 12, 17, 18} use of peripheral blood as the stem cell source, low serum IgG,¹⁹ respiratory viral infection within the first 100 days post-transplant,²⁰ and pulmonary dysfunction before HCT.⁶ Factors that are associated with a poor prognosis once BOS is diagnosed include low serum IgG,¹² early onset after transplantation,^{11, 13} and lack of response to therapy.^{11, 12} However, none of these factors has been consistently reported in the available literature, which is likely constrained by the rarity of this diagnosis.

Restrictive pulmonary dysfunction is associated with, but not diagnostic of chronic GVHD. This finding is often observed in patients with cryptogenic organizing pneumonia (COP), previously called bronchiolitis obliterans organizing pneumonia (BOOP). Restrictive lung dysfunction can have both intra-pulmonary²¹ and extrapulmonary etiologies, including subcutaneous sclerosis of the torso.²²

Measurement of DLCO is frequently done, but not associated with outcomes in patients with chronic GVHD.²³ This measure has the lowest reproducibility, and varies significantly between assessments due to imprecision in measurements. Several reports have demonstrated that DLCO often decreases after HCT, yet can improve over time.^{2, 3}

Data regarding the effect of non-infectious pulmonary complications on survival have been inconsistent. Some studies do not demonstrate any effect on survival.^{5, 24} Other studies clearly demonstrate a lower overall survival in patients with non-infectious pulmonary complications.²⁵ Bronchiolitis obliterans syndrome has been associated with dismal outcomes, with 44% survival at 2 years and 13% survival at 5 years.¹⁰ Even modest progressive airflow obstruction, defined as an annualized decrease of at least 5% per year, has been associated with attributable mortality rates of 9% at 3 years, 12% at 5 years, and 18% at 10 years after transplant. Among patients with chronic GVHD, attributable mortality rates were even higher: 22% at 3 years, 27% at 5 years, and 40% at 10 years.²⁶

In 2005, the NIH held a consensus conference to improve methods of diagnosis and response assessment in chronic GVHD. In this conference, standardized definitions were recommended for BOS: 1) FEV1/FVC ratio of <0.70; 2) FEV1 <75% predicted; 3) air-trapping demonstrated by RV >120% predicted or high resolution computed tomography (HRCT) scan; and 4) absence of an infectious etiology.⁹ A modification was proposed to the criteria in 2010, removing the requirement for a demonstration of air trapping, but specifying that the FEV1 should be <75% predicted or at least 10% lower as compared to pre-transplant PFTs, along with a FEV1/SVC (slow vital capacity) <0.70.¹⁰

Using longitudinal data collected as part of a multicenter, observational study, we tested the pulmonary measures recommended by the 2005 Consensus Conference on Chronic GVHD, to determine their association with non-relapse mortality, survival and patient-reported outcomes.

Methods

Chronic GVHD Consortium: Description of the study cohort

Data are derived from the Chronic GVHD Consortium, a prospective, multicenter, observational study. The protocol was approved by the Institutional Review Board at each site, and all subjects provided written informed consent. Participants were allogeneic HCT recipients at least 2 years of age with chronic GVHD requiring systemic immunosuppressive therapy. Both classic chronic and overlap syndrome were eligible. Cases were classified as incident (enrollment less than 3 months after chronic GVHD diagnosis) or prevalent (enrollment three or more months but less than three years after transplantation). Participants were identified from the population of patients receiving their follow up care at the transplant centers, which is a subset of all patients transplanted by the center. Primary disease relapse, inability to comply with study procedures, and anticipated survival of less than 6 months were exclusion criteria. At enrollment and every 6 months thereafter, clinicians and patients reported standardized information summarizing chronic GVHD organ involvement and symptoms. Incident cases had an additional assessment time point at 3 months after enrollment. Objective medical data including ancillary testing and laboratory results, medical complications, and medication profiles were abstracted through standardized chart review after each visit.

Pulmonary variables

Pulmonary function testing is recommended by the consensus conference, and results associated with each study visit +/- 1 month were recorded when available. Although PFTs were recommended at three month intervals, they were not required. The NIH lung scoring system has two parts. One is a clinical lung symptom score based on symptoms, which will be referred to hereafter as "NIH symptom-based lung score", with Score 0 (no symptoms), Score 1 (shortness of breath with stairs), Score 2 (shortness of breath on flat ground), and Score 3 (shortness of breath at rest or requiring oxygen). The second measure is based on the

lung function score (LFS) calculated from the FEV1 and DLCO corrected for hemoglobin but not alveolar volume.⁹ This score will be called the “NIH PFT-based lung score” to distinguish it from the symptom-based score. The FEV1 and DLCO are converted to a numeric score as follows: >80% = 1; 70–79% = 2; 60–69% = 3; 50–59% = 4; 40–49% = 5; <40% = 6. The LFS = FEV1 score + DLCO score, with a possible range of 2–12 with higher numbers indicating worse dysfunction. The NIH PFT-based Lung Score (0–3) is derived as follows: 0 = FEV1 >80% OR LFS 2, 1 = FEV1 60–79% OR LFS 3–5, 2 = FEV1 40–59% OR LFS 6–9, 3 = FEV1 <39% OR LFS 10–12. In addition, we administered a portable spirometry test during study visits with a hand-held spirometer which records FEV1. The average of 3 attempts was used in the analysis.

Statistical Analyses

We initially performed an unbiased approach on all the measured factors using both univariable and multivariable analysis. Analyses included both cross sectional values and change between assessments. Results were inconsistent when evaluated from a specific time point (such as enrollment or 6 months) or as a kinetic measurement of change over 6 months (data not shown), perhaps because of collinearity between measures. Therefore, we pursued hypothesis-driven analyses instead.

We focused on a set of hypothesized associations between pulmonary measures and non-relapse mortality (NRM), overall survival (OS), and patient-reported outcomes (PROs). The seven measures of interest were: (1) Obstructive lung disease based on PFTs, defined as a FEV1/FVC ratio of <0.7. Two levels of FEV1 were tested: <50% (severe obstructive disease) and 50–80% of predicted (mild and moderate obstructive disease); (2) Restrictive lung disease defined as FVC <80% AND FEV1/FVC <0.7; (3) NIH PFT-based lung score (0–3) (4) NIH symptom-based lung score (0–3); (5) Clinical diagnosis of BOS, as reported by the clinician in the provider survey; (6) Decrease in FEV1 or FVC percent predicted by >10% compared to the first set of PFTs tested after enrollment; and (7) Worsening in NIH symptom-based lung score by one point or greater compared to the first recorded score.

Overall survival was defined from time of enrollment, with patients censored at date of last known to be alive. Non-relapse mortality was defined as death without prior relapse. Cox regression models were fit for OS and NRM using a time-varying covariate model for lung function measures, adjusting for patient characteristics and chronic GVHD global severity calculated without the lung component. Patient characteristics included: platelet count (<100K, 100K), bilirubin (<2 mg/dL, >2 mg/dL), Karnofsky performance score (<80, 80, missing), conditioning regimen (myeloablative, reduced intensity/non-myeloablative), GVHD type (overlap, classic), and HCT comorbidity scale without the lung component.²⁷ We also looked at other covariates, but they were not adjusted due to lack of association with OS or NRM, including: study site (FHCRC, other), case type (incident, prevalent), time from transplant to enrollment (<12 months, ≥12 months), patient age at transplant (<50 years, ≥50 years), donor match (matched related, matched unrelated, mismatched), donor patient gender combination (female into male, other), and prior acute GVHD (yes, no). These covariates considered were chosen a priori based on associations with OS or NRM in previous studies.

In separate models limited to patients with at least two sets of PFTs during the study, we compared survival of patients whose percent predicted FEV1 or FVC declined by 10% or more from the first PFTs recorded after enrollment using time-varying indicators compared to those with stable or improved PFTs. We repeated this model with the FEV1 derived from the hand-held spirometer. We conducted a similar analysis for worsening of one point or greater in NIH symptom-based lung score compared with the first recorded score.

To graphically explore whether change in FEV1 as measured by hand-held spirometry was associated with survival status, we separately plotted available data for patients who ultimately died versus those who were surviving at last follow-up. Data were separately smoothed by a penalized B-spline curve to show the overall trend.

Agreement between FEV1 measured by hand held spirometry and PFTs was graphically displayed with scatter plots and Bland-Altman plots,²⁸ and summarized by the concordance correlation coefficient, using the SAS %CCC macro.^{29, 30} This analysis was done only with data from patients at Fred Hutchinson Cancer Research Center since they were most likely to have the PFTs and hand-held spirometry on the same day.

Utilizing all visit data, we also evaluated the association of each lung measure with patient reported outcomes (PROs), including the Lee symptom scale (lung subscale and overall scale)³¹ and quality of life (SF36-physical component score³² and FACT-BMT³³ trial outcome index) measures as concurrent correlates. Multivariable linear mixed models with random patient effects were used to account for within-patient correlation. All models were adjusted for significant covariates, including: time from transplant to enrollment (<12 months, 12 months), Karnofsky performance score (<80, 80, missing), platelet count (<100K, 100K), NIH global severity (less than mild/mild, moderate, severe), and GVHD type (overlap, classic). A p-value of <0.01 was considered significant because of multiple testing.

Statistical analyses were performed using SAS/STAT software, version 9.3 (SAS Institute, Inc., Cary, NC) and R version 2.15.2 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient characteristics

This analysis included 496 patients who were studied during a total of 1591 visits. Approximately half the patients were assessed at Fred Hutchinson Cancer Research Center. Table 1 shows the characteristics of the patient population. Median follow up time from study entry for survivors was 19.8 months (range 0.3 – 47.7). Two-year overall survival was 81%, and the median survival has not been reached in this population. Among the 496 patients, 166 (33%) had no PFT, 134 (27%) had one set of PFTs, and 196 (40%) had more than one set of PFTs. Patients with at least one set of PFTs (n=330) had similar baseline NIH symptom-based lung scores (p=0.38), OS (HR=0.9, p=0.71), and NRM (HR=0.9, p=0.70) compared to those with no PFTs (n=166), but more severe global scores at baseline (p=0.003) suggesting ascertainment bias in post-transplant testing. Of 330 patients who had at least one set of PFTs, repeated PFTs were not associated with baseline NIH symptom-based lung score (p=0.89) or baseline global severity (p=0.38). Pre-transplant PFT information was missing for 50 (10%) of patients, often because patients were enrolled at a different center than where they were transplanted, while PFTs were missing for patients at diagnosis of chronic GVHD in 254 (51%) and at enrollment onto the study in 237 (48%) of patients. Overall, 845 (53%) of all visits had recorded PFTs.

Pulmonary dysfunction was present at enrollment in 34% of patients based on PFTs and 25% based on symptoms (Table 1). Using all assessments, mild-moderate obstructive physiology (FEV1 50–80%, FEV1/FVC <0.70) was identified in 137 visits (16%), and 54 visits (6%) had severe obstruction with FEV1 <50%. BOS was reported at 122 visits from 59 patients (32 patients at enrollment). (Table 2)

Association of cross-sectional pulmonary measures with mortality

We tested whether cross-sectional pulmonary measures were associated with mortality, to see if any isolated assessment is associated with subsequent death. The NIH symptom-based lung score was associated with OS and NRM in the multivariable Cox regression models (Table 2). For example, compared to no lung symptoms, a NIH symptom-based lung score of 3 (shortness of breath at rest or requiring oxygen) was associated with higher NRM (HR 5.6, 95% CI 1.3–17.3, $p=0.01$) in the multivariable models. Even a NIH symptom-based lung score of 1 was associated with worse OS (HR 2.0, 95% CI 1.2–3.3, $p=0.005$). These findings were independent of Karnofsky performance status, and the interaction effect was not statistically significant ($p=0.34$ in NRM and $p=0.10$ in OS models), although the two measures had moderate correlation with each other (Pearson correlation = -0.34). In patients with a NIH symptom-based lung score of 2 or 3 at enrollment, median OS was 36 months and survival was 90% at 6 months, 90% at 12 months and 82% at 18 months, suggesting a slow rather than rapid rate of mortality (Figure 1). None of the other measures including obstructive or restrictive PFTs, the score based solely on the PFTs, or the clinician's indicator of BOS were significantly associated with OS or NRM.

FEV1 as assessed by hand-held spirometry and PFTs had good agreement with a concordance correlation coefficient between the two measures of 0.74 (95% CI: 0.71–0.77) (Figure 2). Hand-held spirometry values for FEV1 were available for 1410 visits (89%) and trends for association with NRM and OS were observed but were not statistically significant ($p=0.05$) at the predefined cutoffs of $<80\%$ and $<50\%$.

Association of changes in lung measures with mortality

We tested whether serial pulmonary measures were associated with mortality, to see if changes in measures are associated with subsequent death. Using all available data from PFTs ($n=515$) or hand-held spirometry ($n=934$) in models where change from the first recorded measure was considered a time-varying covariate, we found no association between $>10\%$ change in percent predicted FEV1 and either NRM or OS. When hand-held spirometry results were analyzed graphically, patients who died during the observational period had a decrease in their FEV1 compared to patients who survived (Figure 3). The spaghetti plot shows all available data for the two groups with the penalized B spline demonstrating the overall trend.

Using all available data from the NIH symptom-based lung score ($n=1089$) in a model where change from the first recorded score was considered a time-varying covariate, worsening at a visit by at least one point ($n=202$) was associated with both higher NRM (HR 3.9, 95% CI 1.8–8.6, $p=0.001$) and worse overall survival (HR 2.9, 95% CI 1.5–5.5, $p=0.001$) compared to visits where patients had stable lung symptom scores ($n=759$).

Association of pulmonary measures with self-reported outcomes

In multivariable models, the NIH symptom-based lung score was also highly correlated with patient-reported outcomes including the Lee lung symptom score, Lee summary symptom score, SF36-physical component score, and FACT-BMT trial outcome index (all $p<0.001$). (Table 3) Obstructive PFTs and bronchiolitis obliterans were associated with the Lee lung symptom score. The SF36-mental component was not associated with any lung function measures (data not shown).

Discussion

These results indicate that the NIH symptom-based lung score is associated with OS, NRM, and PRO measures when measured at a single time point, and worsening of the score over

time is associated with NRM and OS. The performance of the NIH symptom-based lung score is encouraging as it is relatively easy to capture and is not dependent on a test that is subject to operator error and engenders additional costs. Its association with outcome is independent of Karnofsky performance status and other clinical factors. Additionally, although performance of a NIH symptom-based lung score of 3 (shortness of breath at rest or requiring oxygen) intuitively is associated with worse outcomes, it is notable that patients who have a NIH symptom-based lung score of 1 (shortness of breath with stairs) also had a worse outcome compared to those with a score of 0.

PFT measurements at any cross-sectional time point or worsening compared with values at the first recorded measure were not associated with OS and NRM. This is surprising, as most studies demonstrate that the rate of change in PFTs predicts survival even when single assessments do not. Chien et al. demonstrated that even a modest reduction in FEV1 from pre-transplant to 1 year post transplant was associated with an inferior survival,²⁶ and Dudek et al¹² reported that improvement or stability in PFTs was correlated with improved survival, and worsening of the PFTs correlated with a significantly worse survival.¹² In the lung transplant literature, both severity at onset and rate of decline were associated with OS and NRM.³⁴ Four factors may have contributed to the lack of correlation between decline in PFT parameters and OS and NRM in our study. First, patients were followed for a relatively short period of time, perhaps not long enough to reveal the serial declines that would be predictive of survival. This is compounded by the fact that the PFT parameters evaluated may fluctuate up to 8% predicted as part of normal variation, and also will be altered in the setting of infection or COP. Thus, these parameters are not robust in sensitivity. Secondly, our population is limited due to the relatively few PFTs obtained on the cohort, notably 33% had no PFTs and 27% only had one PFT available. Thus 60% of the total cohort could not be evaluated for serial declines. Third, there were only a small number of patients with severe obstruction in the cohort, limiting our ability to understand the relative impact of severe obstruction on OS and NRM. Finally, the background mortality rate from non-pulmonary causes may be higher in our chronic GVHD population than in others where pulmonary dysfunction is the major life-threatening condition. Given that our data reveal a direct association of symptomatic lung dysfunction with OS and NRM in patients with chronic GVHD, more consistent PFT monitoring could be of potential benefit to capture these patients at an earlier stage of disease and potentially initiate interventions with a greater likelihood of success.

Clinician documentation of BOS did not have a significant association with OS and NRM, although BOS showed some correlation with the Lee lung symptom score. In this cohort, BOS was identified by clinicians in only 32 (6%) patients at enrollment and 59 (14%) patients at any time during the study (122 visits). Although we do not have the detailed information needed to apply the older NIH criteria for diagnosis of BOS, if we consider PFT data alone as a surrogate for the diagnosis of BOS using the modified chronic GVHD diagnostic criteria, 6% of patients had evidence of BOS at enrollment.

Survival for patients with NIH symptom-based lung scores worse than 1 was better than expected compared to historic controls. This difference may be related to the populations studied, the relatively short period of follow up, improvements in supportive care, or possibly exaggerated diagnosis (due to lack of precedent PFTs and possible long-standing lung findings that are unassociated with HCT). It is also possible that patients with pulmonary dysfunction have improved outcomes due to increased awareness, although no early interventions are reported as highly effective. Although clinicians are not always consistent in checking PFTs, they may be more aware of potential pulmonary dysfunction associated with chronic GVHD, therefore identifying a higher proportion of patients and

starting therapy earlier in the natural history of the disease. This hypothesis is supported by the finding that patients with severe chronic GVHD were more likely to have PFTs done.

Missing data limited the power of our analysis. Approximately half of visits were missing PFT results, suggesting that despite recommendations to follow PFTs frequently in patients with chronic GVHD, clinicians do not consistently follow this recommendation, which has been observed in other analyses as well.²³ Another possible explanation for the missing PFTs is that we captured only test results that were within one month of a study visit. It is possible that many PFTs were performed outside this window. Somewhat mitigating the concern about loss of power due to missing PFTs is the fact that we had hand-held spirometry data on the majority of the patients and did not find a consistent association of FEV1 with OS and NRM. However, hand-held spirometry measurements in patients who died during the observation period showed a trend for decreased volumes. While these data provide insight into this otherwise absent cohort, these data would suggest that hand-held spirometry, while useful, is not a replacement for PFTs.

Even though results from a single PFT and changes over time did not predict OS and NRM, we believe that PFTs should still be performed because our analyses addressed only the prognostic significance of PFTs, not whether they have a role in clarifying diagnoses or guiding management of patients with chronic GVHD. If PFTs can detect changes before symptoms develop, then treatments can be instituted earlier, with the hope that morbidity and mortality can be decreased. One study suggested that a brief course of therapy at the onset of pulmonary dysfunction may lead to improved outcomes.³⁵ The question of whether BOS-directed therapy can improve prognosis in patients with this diagnosis is also being addressed in an ongoing multicenter study that uses prednisone, fluticasone, montelukast, and azithromycin to treat patients with BOS who are within 6 months of diagnosis (NCI: NCT01307462). This study will assess the relative impact of early, lung-directed treatment compared to historical controls. This study could also provide an improved understanding of the benefit of early improvement in PFT function with treatment. Future studies that follow serial PFTs in patients with new onset obstructive disease would help provide insight to the questions about the prognostic impact of PFT abnormalities in lung disease associated with chronic GVHD.

In summary, there is strong evidence that the NIH symptom-based lung score, either used cross-sectionally or as a serial measure over time, is statistically associated with NRM, OS, and PRO, and should be ascertained in patients with chronic GVHD.

Acknowledgments

This work was supported by grants CA118953 and CA163438 from the National Institutes of Health (NIH). The Chronic GVHD Consortium (U54 CA163438) is a part of the NIH Rare Diseases Clinical Research Network (RDCRN), supported through collaboration between the NIH Office of Rare Diseases Research (ORDR) at the National Center for Advancing Translational Science (NCATS), the National Cancer Institute, and the Fred Hutchinson Cancer Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Chien JW, Martin PJ, Flowers ME, Nichols WG, Clark JG. Implications of early airflow decline after myeloablative allogeneic stem cell transplantation. *Bone Marrow Transplantation*. 2004; 33:759–764. [PubMed: 14968136]
2. Gore EM, Lawton CA, Ash RC, Lipchik RJ. Pulmonary function changes in long-term survivors of bone marrow transplantation. *International Journal of Radiation Oncology*Biology*Physics*. 1996; 36:67–75.

3. Marras TK, Chan CK, Lipton JH, Messner HA, Szalai JP, Laupacis A. Long-term pulmonary function abnormalities and survival after allogeneic marrow transplantation. *Bone Marrow Transplant.* 2004; 33:509–517. [PubMed: 14716347]
4. Ralph DD, Springmeyer SC, Sullivan KM, Hackman RC, Storb R, Thomas ED. Rapidly progressive air-flow obstruction in marrow transplant recipients. Possible association between obliterative bronchiolitis and chronic graft-versus-host disease. *Am Rev Respir Dis.* 1984; 129:641–644. [PubMed: 6370061]
5. Sakaida E, Nakaseko C, Harima A, et al. Late-onset noninfectious pulmonary complications after allogeneic stem cell transplantation are significantly associated with chronic graft-versus-host disease and with the graft-versus-leukemia effect. *Blood.* 2003; 102:4236–4242. [PubMed: 12907447]
6. Savani BN, Montero A, Srinivasan R, et al. Chronic GVHD and Pretransplantation Abnormalities in Pulmonary Function Are the Main Determinants Predicting Worsening Pulmonary Function in Long-term Survivors after Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation.* 2006; 12:1261–1269. [PubMed: 17162207]
7. Tait RC, Burnett AK, Robertson AG, et al. Subclinical pulmonary function defects following autologous and allogeneic bone marrow transplantation: Relationship to total body irradiation and graft-versus-host disease. *International Journal of Radiation Oncology*Biophysics*Physics.* 1991; 20:1219–1227.
8. Yoshihara S, Yanik G, Cooke KR, Mineishi S. Bronchiolitis Obliterans Syndrome (BOS), Bronchiolitis Obliterans Organizing Pneumonia (BOOP), and Other Late-Onset Noninfectious Pulmonary Complications following Allogeneic Hematopoietic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation.* 2007; 13:749–759. [PubMed: 17580252]
9. 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. *Biology of Blood and Marrow Transplantation.* 2005; 11:945–956. [PubMed: 16338616]
10. Chien JW, Duncan S, Williams KM, Pavletic SZ. Bronchiolitis Obliterans Syndrome After Allogeneic Hematopoietic Stem Cell Transplantation—An Increasingly Recognized Manifestation of Chronic Graft-versus-Host Disease. *Biology of Blood and Marrow Transplantation.* 2010; 16:S106–S114. [PubMed: 19896545]
11. Au BKC, Au MA, Chien JW. Bronchiolitis Obliterans Syndrome Epidemiology after Allogeneic Hematopoietic Cell Transplantation. *Biology of Blood and Marrow Transplantation.* 2011; 17:1072–1078. [PubMed: 21126596]
12. Dudek AZ, Mahaseth H, DeFor TE, Weisdorf DJ. Bronchiolitis obliterans in chronic graft-versus-host disease: analysis of risk factors and treatment outcomes. *Biology of Blood and Marrow Transplantation.* 2003; 9:657–666. [PubMed: 14569562]
13. Bergeron A, Godet C, Chevret S, et al. Bronchiolitis obliterans syndrome after allogeneic hematopoietic SCT: phenotypes and prognosis. *Bone Marrow Transplant.* 2012
14. Curtis DJ, Smale A, Thien F, Schwarer AP, Szer J. Chronic airflow obstruction in long-term survivors of allogeneic bone marrow transplantation. *Bone Marrow Transplant.* 1995; 16:169–173. [PubMed: 7581118]
15. Ho VT, Weller E, Lee SJ, Alyea EP, Antin JH, Soiffer RJ. Prognostic factors for early severe pulmonary complications after hematopoietic stem cell transplantation. *Biology of Blood and Marrow Transplantation.* 2001; 7:223–229. [PubMed: 11349809]
16. Schwarer AP, Hughes JMB, Trotman-Dickenson B, Krausz T, Goldman JM. A chronic pulmonary syndrome associated with graft-versus-host disease after allogeneic marrow transplantation. *Transplantation.* 1992; 54:1002–1008. [PubMed: 1465767]
17. Ringdén O, Remberger M, Ruutu T, et al. Increased Risk of Chronic Graft-Versus-Host Disease, Obstructive Bronchiolitis, and Alopecia With Busulfan Versus Total Body Irradiation: Long-Term Results of a Randomized Trial in Allogeneic Marrow Recipients With Leukemia. *Blood.* 1999; 93:2196–2201. [PubMed: 10090927]
18. Clark JG, Crawford SW, Madtes DK, Sullivan KM. Obstructive Lung Disease after Allogeneic Marrow Transplantation. *Annals of Internal Medicine.* 1989; 111:368–376. [PubMed: 2669592]

19. Holland HK, Wingard JR, Beschoner WE, Saral R, Santos GW. Bronchiolitis obliterans in bone marrow transplantation and its relationship to chronic graft-v-host disease and low serum IgG. *Blood*. 1988; 72:621–627. [PubMed: 3042044]
20. Erard V, Chien JW, Kim HW, et al. Airflow Decline after Myeloablative Allogeneic Hematopoietic Cell Transplantation: The Role of Community Respiratory Viruses. *Journal of Infectious Diseases*. 2006; 193:1619–1625. [PubMed: 16703503]
21. Wolff D, Reichenberger F, Steiner B, et al. Progressive interstitial fibrosis of the lung in sclerodermoid chronic graft-versus-host disease. *Bone Marrow Transplantation*. 2002; 29:357–360. [PubMed: 11896434]
22. Wingard JR, Vogelsang GB, Deeg HJ. Stem Cell Transplantation: Supportive Care and Long-Term Complications. *ASH Education Program Book*. 2002; 2002:422–444.
23. Hildebrandt GC, Fazekas T, Lawitschka A, et al. Diagnosis and treatment of pulmonary chronic GVHD: report from the consensus conference on clinical practice in chronic GVHD. *Bone Marrow Transplant*. 2011; 46:1283–1295. [PubMed: 21441964]
24. Duncker C, Dohr D, Harsdorf S, et al. Non-infectious lung complications are closely associated with chronic graft-versus-host disease: a single center study of incidence, risk factors and outcome. *Bone Marrow Transplant*. 2000; 25:1263–1268. [PubMed: 10871731]
25. Nishio N, Yagasaki H, Takahashi Y, et al. Late-onset non-infectious pulmonary complications following allogeneic hematopoietic stem cell transplantation in children. *Bone Marrow Transplant*. 2009; 44:303–308. [PubMed: 19349954]
26. Chien JW, Martin PJ, Gooley TA, et al. Airflow Obstruction after Myeloablative Allogeneic Hematopoietic Stem Cell Transplantation. *Am J Respir Crit Care Med*. 2003; 168:208–214. [PubMed: 12649126]
27. Sorror ML, Maris MB, Storb R, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood*. 2005; 106:2912–2919. [PubMed: 15994282]
28. Martin Bland J, Altman D. Statistical Methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986; 327:307–310.
29. Crawford SB, Kosinski AS, Lin H, Williamson JM, Barnhart HX. Computer programs for the concordance correlation coefficient. *Comput Methods Programs Biomed*. 2007; 88:62–74. [PubMed: 17709153]
30. Barnhart HX, Haber M, Song J. Overall Concordance Correlation Coefficient for Evaluating Agreement Among Multiple Observers. *Biometrics*. 2002; 58:1020–1027. [PubMed: 12495158]
31. Lee SJ, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biology of Blood and Marrow Transplantation*. 2002; 8:444–452. [PubMed: 12234170]
32. McHorney CA, Ware JEJ, 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; 31:247–263. [PubMed: 8450681]
33. 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; 19:357–368. [PubMed: 9051246]
34. Finlen Copeland CA, Snyder LD, Zaas DW, Turbyfill WJ, Davis WA, Palmer SM. Survival After Bronchiolitis Obliterans Syndrome Among Bilateral Lung Transplant Recipients. *Am J Respir Crit Care Med*. 2010
35. Sanchez J, Torres A, Serrano J, et al. Long-term follow-up of immunosuppressive treatment for obstructive airways disease after allogeneic bone marrow transplantation. *Bone Marrow Transplant*. 1997; 20:403–408. [PubMed: 9339757]

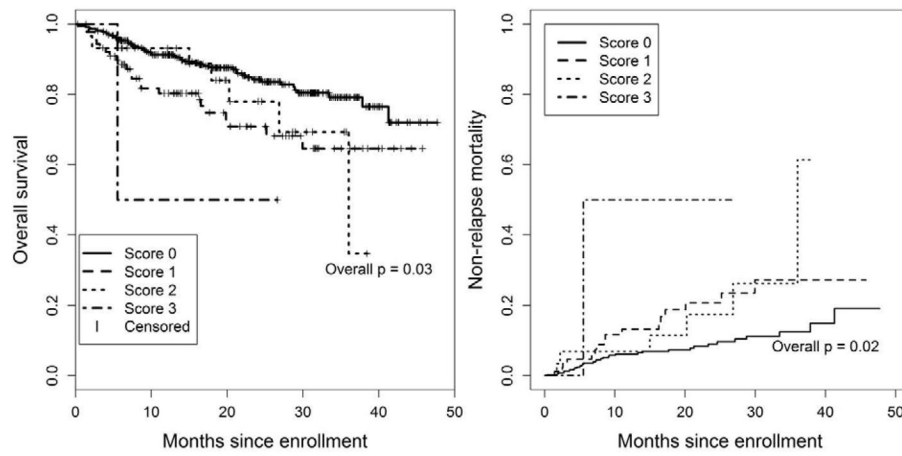


Figure 1. Kaplan-Meier overall survival and cumulative incidence of non-relapse mortality, according to NIH symptom-based lung score at enrollment

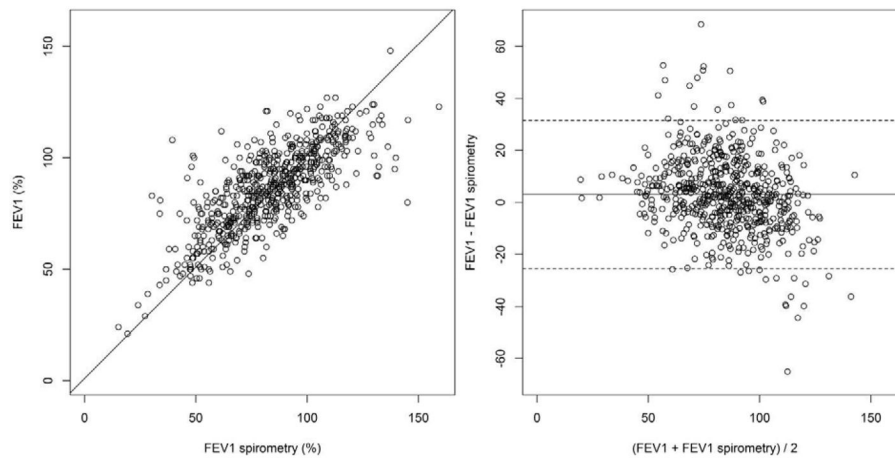


Figure 2. Scatter plot (left) and Bland-Altman plot (right) between hand-held spirometry and FEV1 on PFTs. On the Bland-Altman plot, the mean difference between two measures was centered around zero (solid line), with the 95% limits of agreement shown by dashed lines.

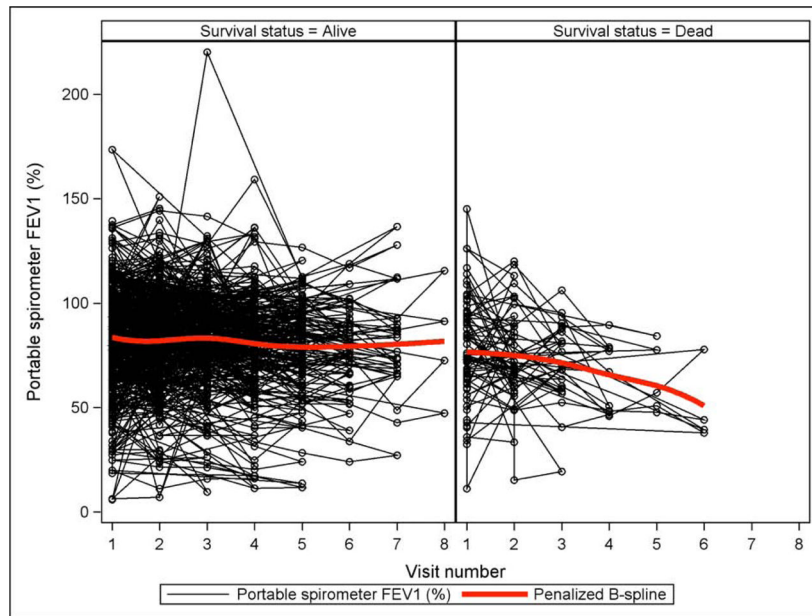


Figure 3. Spaghetti plot of hand-held spirometry FEV1 according to survival status

Table 1

Characteristics of the Study Population at enrollment (n=496)

Characteristics	Category	n	Count (%)
Case type	Incident	496	281 (57%)
	Prevalent		215 (43%)
Adult or child	Adult (18+)	496	482 (97%)
	Child (2–17)		14 (3%)
Patient gender	Female	496	206 (42%)
	Male		290 (58%)
Diagnosis	Acute Myeloid Leukemia	496	164 (33%)
	Acute Lymphoblastic Leukemia		63 (13%)
	Chronic Myeloid Leukemia		27 (5%)
	Chronic Lymphocytic Leukemia		38 (8%)
	Myelodysplastic Syndrome		73 (15%)
	Non-Hodgkin Lymphoma		70 (14%)
	Hodgkin Lymphoma		17 (3%)
	Multiple Myeloma		22 (4%)
	Aplastic Anemia		6 (1%)
	Other		16 (3%)
Disease status	Early	495	164 (33%)
	Intermediate		214 (43%)
	Advanced		117 (24%)
Transplant source	Bone marrow	496	35 (7%)
	Cord blood		23 (5%)
	Peripheral blood		438 (88%)
Transplant type	Myeloablative	495	250 (51%)
	Not myeloablative		245 (49%)
Total Body Irradiation	No TBI	496	178 (36%)
	TBI myeloablative		141 (28%)
	TBI reduced intensity/non-myeloablative		177 (36%)
Patient CMV status	Negative	493	214 (43%)
	Positive		279 (57%)
Donor CMV status	Negative	489	299 (61%)
	Positive		190 (39%)
Donor match	Matched related	495	216 (44%)
	Matched unrelated		200 (40%)
	Mismatched		79 (16%)

Characteristics	Category	n	Count (%)
Site	Fred Hutchinson Cancer Research Center	496	227 (46%)
	Moffitt Cancer Center		13 (3%)
	University of Minnesota		56 (11%)
	Dana-Faber Cancer institute		59 (12%)
	Stanford University Medical Center		69 (14%)
	Northwest Childrens Hospital		13 (3%)
	Vanderbilt University Medical Center		40 (8%)
	Medical College of Wisconsin		16 (3%)
	Washington University Medical Center		3 (1%)
Prior grade acute GVHD	Grade 0-I	496	229 (46%)
	Grade II-IV		267 (54%)
NIH global severity score	Less than mild	496	3 (1%)
	Mild		41 (8%)
	Moderate		261 (53%)
	Severe		191 (38%)
Clinical lung category based on interpretation of PFTs	Normal	259	170 (66%)
	Suggestive of restrictive dysfunction		44 (17%)
	Mild to moderate obstruction		34 (13%)
	Severe obstruction		11 (4%)
NIH symptom-based lung score (0-3)	Score 0 (no symptoms)	496	375 (75%)
	Score 1 (shortness of breath with stairs)		88 (18%)
	Score 2 (shortness of breath on flat ground)		30 (6%)
	Score 3 (shortness of breath at rest or requiring oxygen)		3 (1%)
NIH PFT-based lung score (0-3)	Score 0 (FEV1 >80% or LFS 2)	261	59 (23%)
	Score 1 (FEV1 60-79% or LFS 3-5)		141 (54%)
	Score 2 (FEV1 40-59% or LFS 6-9)		53 (20%)
	Score 3 (FEV1 <39% or LFS 10-12)		8 (3%)

Abbreviations: GVHD, graft-versus-host disease; NIH, National Institutes of Health; PFTs, pulmonary function tests

Table 2

Association of pulmonary measures, non-relapse mortality and overall survival from multivariable time-varying Cox regression models.

Measure	Definition	N (%)	Non-relapse mortality			Overall Survival		
			HR	95% CI	p-value	HR	95% CI	p-value
Lung category (n=845)	Normal	520 (62%)	1.0		0.53*	1.0		0.55*
	Suggestive of restrictive	134 (16%)	1.5	(0.6-3.5)	0.33	1.3	(0.6-2.6)	0.45
	Mild and moderate obstructive	137 (16%)	0.6	(0.1-1.8)	0.43	0.7	(0.2-1.7)	0.45
	Severe obstructive	54 (6%)	1.1	(0.2-3.9)	0.85	1.5	(0.5-3.8)	0.47
NIH PFT-based lung score (0-3) (n=853) [†]	Score 0	213 (25%)	1.0		0.13*	1.0		0.42*
	Score 1	421 (49%)	2.2	(0.7-9.6)	0.23	1.4	(0.6-3.6)	0.41
	Score 2	180 (21%)	3.9	(1.2-17.5)	0.04	2.0	(0.8-5.4)	0.13
	Score 3	39 (5%)	1.9	(0.2-12.2)	0.49	2.0	(0.5-7.0)	0.27
NIH symptom-based lung score (0-3) (n=1585)	Score 0 (no symptoms)	1163 (73%)	1.0		0.02*	1.0		0.02*
	Score 1 (shortness of breath with stairs)	304 (19%)	2.0	(1.1-3.8)	0.03	2.0	(1.2-3.3)	0.005
	Score 2 (shortness of breath on flat ground)	96 (6%)	2.3	(0.9-5.2)	0.06	1.8	(0.8-3.6)	0.12
	Score 3 (shortness of breath at rest or requiring oxygen)	22 (1%)	5.6	(1.3-17.3)	0.01	3.6	(0.8-10.4)	0.04
Clinical diagnosis of bronchiolitis obliterans syndrome (n=1583)		122 (8%)	1.7	(0.7-3.5)	0.21	1.5	(0.7-2.8)	0.21

* Overall p-value

[†] NIH PFT-based lung score: 0 = FEV1 >80% or LFS 2, 1 = FEV1 60-79% OR LFS 3-5, 2 = FEV1 40-59% OR LFS 6-9, 3 = FEV1 39% OR LFS 10-12

Table 3

Association of pulmonary measures with patient-reported symptoms and quality of life from multivariable linear mixed models.

Measure	Definition	Lee Lung Symptom Score			Lee Summary Symptom Score			SF-36 PCS			FACT-BMT TOI		
		n	Estimate	p-value	n	Estimate	p-value	n	Estimate	p-value	n	Estimate	p-value
Lung category (n=845)	Normal	673	0 (reference)	<.001*	676	0 (reference)	0.47*	632	0 (reference)	0.05*	640	0 (reference)	0.47*
	Suggestive of restrictive		2.8	0.03		1.4	0.26		-2.5	0.01		-1.8	0.20
	Mild and moderate obstructive		5.8	<.001		1.3	0.30		-1.6	0.10		-0.8	0.55
	Severe obstructive		7.3	<.001		-0.5	0.79		0.01	0.99		1.2	0.62
NIH PFT-based lung score (0-3) (n=853)	Score 0	680	0 (reference)	0.03*	683	0 (reference)	0.14*	638	0 (reference)	0.46*	646	0 (reference)	0.88*
	Score 1		1.3	0.22		-0.5	0.58		-0.5	0.57		0.1	0.93
	Score 2		4.8	0.003		-2.9	0.05		0.3	0.80		0.6	0.71
	Score 3		2.9	0.24		-4.5	0.04		-2.5	0.24		2.4	0.42
NIH symptom-based lung score (0-3) (n=1585)	Score 0 (no symptoms)	1309	0 (reference)	<.001*	1312	0 (reference)	<.001*	1234	0 (reference)	<.001*	1254	0 (reference)	<.001*
	Score 1 (shortness of breath with stairs)		5.5	<.001		3.0	<.001		-2.9	<.001		-3.4	<.001
	Score 2 (shortness of breath on flat ground)		12.3	<.001		5.2	<.001		-4.2	<.001		-3.2	0.01
	Score 3 (shortness of breath at rest or requiring oxygen)		21.4	<.001		5.5	0.02		-7.2	0.002		-5.3	0.09
Clinical diagnosis of bronchiolitis obliterans syndrome (n=1583)		1307	6.8	<.001	1310	1.7	0.13	1232	-2.0	0.04	1252	-0.6	0.66

Abbreviations: PFTs, pulmonary function tests; SF36 PCS, Medical Outcomes Study Physical Component Scale; FACT-BMT TOI, Functional Assessment of Cancer Therapy – Bone Marrow Transplant subscale Trial Outcome Index.

* Overall p-value