Biol Blood Marrow Transplant 20 (2014) 337–344

Pulmonary Symptoms Measured by the National Institutes of Health Lung Score Predict Overall Survival, Nonrelapse Mortality, and Patient-Reported Outcomes In Chronic Graft-Versus-Host Disease

Jeanne Palmer [1,*], Kirsten Williams [2], Yoshihiro Inamoto [3], Xiaoyu Chai [3], Paul J. Martin [3], Linus H. Santo Tomas [4], Corey Cutler [5], Daniel Weisdorf [6], Brenda F. Kurland [7], Paul A. Carpenter [3], Joseph Pidala [8], Steven Z. Pavletic [9], William Wood [10], David Jacobsohn [2], Sally Arai [11], Mukta Arora [6], Madan Jagasia [12], Georgia B. Vogelsang [13], Stephanie J. Lee [3]

[1] *Division of Hematology/Oncology, Medical College of Wisconsin, Milwaukee, Wisconsin*
[2] *Division of Blood and Marrow Transplantation, Children's National Medical Center, Center for Cancer and Blood Disorders, Washington, District of Columbia*
[3] *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington*
[4] *Pulmonary Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin*
[5] *Hematologic Malignancies, Dana-Farber Cancer Institute, Boston, Massachusetts*
[6] *Blood and Marrow Transplant Program, University of Minnesota, Minneapolis, Minnesota*
[7] *Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania*
[8] *Blood and Marrow Transplantation, Moffitt Cancer Center, Tampa, Florida*
[9] *National Cancer Institute, National Institutes of Health, Bethesda, Maryland*
[10] *Linenberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina*
[11] *Division of Blood and Marrow Transplantation, Stanford University Medical Center, Stanford, California*
[12] *Hematology and Stem Cell Transplant Program, Vanderbilt University Medical Center, Nashville, Tennessee*
[13] *Oncology, Johns Hopkins Hospital, Baltimore, Maryland*

Article history:
Received 7 September 2013
Accepted 27 November 2013

Key Words:
Chronic graft-versus-host disease
Bronchiolitis obliterans syndrome

A B S T R A C T

The 2005 National Institutes of Health (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 nonrelapse 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 transplantation characteristics and nonlung 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 ($P = .02$), OS ($P = .02$), patient-reported symptoms ($P < .001$) and functional status ($P < .001$). 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 symptom score of 0 to 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 forced expiratory volume in 1 second (FEV1) is less than 80% of expected and FEV1/forced vital capacity (FVC) <.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 >.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 transplantation 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

Financial disclosure: See Acknowledgments on page 343.
* Correspondence and reprint requests: Jeanne Palmer, MD, Blood and Marrow Transplant Program, Hematology Oncology, 5777 East Mayo Blvd, Phoenix, AZ 85054.
E-mail address: palmer.jeanne@mayo.edu (J. Palmer).

1083-8791/$ – see front matter © 2014 American Society for Blood and Marrow Transplantation. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbmt.2013.11.025

prophylaxis [12], 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 transplantation [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, previously called *bronchiolitis obliterans organizing pneumonia*. Restrictive lung dysfunction can have both intrapulmonary [21] 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 because of imprecision in measurements. Several reports have demonstrated that DLCO often decreases after HCT, yet it can improve over time [2,3].

Data regarding the effect of noninfectious 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 noninfectious 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 transplantation. 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 $<.70$; (2) FEV1 $<75\%$ predicted; (3) air-trapping demonstrated by residual volume $>120\%$ predicted or high resolution computed tomography scan; and (4) absence of an infectious etiology [9]. A modification to the criteria was proposed 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 with pretransplantation PFTs, along with a FEV1/slow vital capacity $<.70$ [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 nonrelapse 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. Patients with either classic chronic and overlap syndrome were eligible. Cases were classified as incident (enrollment less than 3 months after chronic GVHD diagnosis) or prevalent (enrollment 3 or more months but less than 3 years after transplantation). Participants were identified from the population of patients receiving their follow-up care at the transplantation centers, which is a subset of all patients who underwent transplantation at 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 $\pm$ 1 month were recorded when available. Although PFTs were recommended at 3-month intervals, they were not required. The NIH lung scoring system has 2 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% to 79% = 2; 60% to 69% = 3; 50% to 59% = 4; 40% to 49% = 5; and $<40\% = 6$. The LFS = FEV1 score + DLCO score, with a possible range of 2 to 12, with higher numbers indicating worse dysfunction. The NIH PFT-based Lung Score (0 to 3) is derived as follows: 0 = FEV1 $>80\%$ or LFS 2; 1 = FEV1 60% to 79% or LFS 3 to 5; 2 = FEV1 40% to 59% or LFS 6 to 9; and 3 = FEV1 $\leq 39\%$ or LFS 10 to 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 nonrelapse mortality (NRM), overall survival (OS), and patient-reported outcomes (PROs). The 7 measures of interest were as follows: (1) obstructive lung disease based on PFTs, defined as a FEV1/FVC ratio of $<.7$ (2 levels of FEV1 were tested: $<50\%$ [severe obstructive disease] and 50% to 80% of predicted [mild and moderate obstructive disease]); (2) restrictive lung disease defined as FVC $\leq 80\%$ AND FEV1/FVC $\geq .7$; (3) NIH PFT-based lung score (0 to 3); (4) NIH symptom-based lung score (0 to 3); (5) clinical diagnosis of BOS, as reported by the clinician in the provider survey; (6) decrease in FEV1 or FVC percent predicted by $\geq 10\%$ compared with the first set of PFTs tested after enrollment; and (7) worsening in NIH symptom-based lung score by 1 point or greater compared with the first recorded score.

Overall survival was defined from time of enrollment, with patients censored at date of last known to be alive. Nonrelapse 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/nonmyeloablative), GVHD type (overlap, classic), and HCT comorbidity scale without the lung component [27]. We also looked at other covariates, although they were not adjusted because of lack of association with OS or NRM, including: study site (Fred Hutchinson Cancer Research Center, other), case type (incident, prevalent), time from transplantation to enrollment (<12 months, ≥ 12 months), patient age at transplantation (<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 2 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 with those with stable or improved PFTs. We repeated this model with the FEV1 derived from the hand-held