Biol Blood Marrow Transplant xxx (2016) 1–6

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

Outcome after Transplantation According to Reduced-Intensity Conditioning Regimen in Patients Undergoing Transplantation for Myelofibrosis

Marie Robin [1,2,*], Raphael Porcher [3,4], Christine Wolschke [5], Flore Sicre de Fontbrune [1,2], Haefaa Alchalby [5], Maximilian Christopeit [5], Bruno Cassinat [5], Tatjana Zabelina [5], Régis Peffault de Latour [1,2], Francis Ayuk [5], Gérard Socié [1,2], Nicolaus Kröger [5]

[1] *Hematology Transplantation, Saint-Louis Hospital, Assistance Publique-Hôpitaux de Paris, 2 University Paris 7, INSERM UMR 1160, Paris, France*
[2] *Université Paris Descartes, Sorbonne Paris Cité, Paris, UMR-S 1153, France*
[3] *INSERM U1153, Paris, France. Assistance Publique-Hôpitaux de Paris, Hôtel-Dieu, Centre d'Épidémiologie Clinique, Paris, France*
[4] *Department of Stem Cell Transplantation, University Medical Center Hamburg, Hamburg, Germany*
[5] *Biologie Cellulaire, Saint-Louis Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France*

Article history:
Received 21 January 2016
Accepted 23 February 2016

Key Words:
Myelofibrosis
Reduced-intensity conditioning regimen

A B S T R A C T

Allogeneic hematopoietic stem cell transplantation remains the sole curative option for myelofibrosis. Many transplantation recipients receive a reduced-intensity conditioning (RIC) regimen owing to age or comorbidities; however, there is little published evidence to guide the choice of RIC regimen. In this study, we compared outcomes in patients who received 1 of 2 frequently used RIC regimens for patients with myelofibrosis: fludarabine-busulfan (FB) and fludarabine-melphalan (FM). A total of 160 patients underwent a RIC allograft procedure (FB group, n = 105; FM group, n = 55). We have developed a complex statistical model involving weighting and adjustment to permit comparison between these 2 groups. After weighting, the incidence of acute graft-versus-host disease (GVHD) was 62% in the FM group and 31% in the FB group ($P = .001$), and the corresponding incidence of chronic GVHD was 49% and 53%, respectively. The 7-year progression-free survival was were 52% in the FM group versus 33% in the FB group, and the 7-year overall survival rate 52% in the FM group versus 59% in the FB group. Nonrelapse mortality (NRM) was 43% in the FM group and 31% in the FB group. Multivariable analyses revealed no significant differences in PFS between the 2 groups; however, the relapse rate was significantly lower in the FM group (hazard ratio, 9.21; $P = .008$), whereas a trend toward reduced NRM was seen in the FB group (hazard ratio, 0.51; $P = .068$). In conclusion, both regimens appear to be efficient in mediating disease control and can be used to successfully condition patients with myelofibrosis. The FM regimen appears to induce more NRM than the FB regimen, but with augmented control of disease, leading to comparable overall survival rates for both regimens.

INTRODUCTION

Despite major advances in our understanding of the pathogenesis of myelofibrosis and the development of targeted therapies, allogeneic hematopoietic stem cell transplantation (HSCT) remains the sole potential curative therapy able to induce total disappearance of marrow fibrosis and all signs of the disease. Unfortunately, HSCT may be followed by failure, secondary to either toxicity or relapse, in up to one-half of patients [1-8]. So-called "reduced-intensity" conditioning (RIC) regimens were developed at the end of the 1990s to overcome high nonrelapse mortality (NRM) rates in patients with hematologic malignancies undergoing myeloablative protocols and to permit extension of HCST as a therapeutic option for older and/or more fragile patients.

RIC regimens are heterogeneous and produce both intermediate myeloablation and immunosuppression, meaning that subsequent autologous reconstitution is possible but not systematic [9]. Among the many available RIC regimens, the most common used in patients with myelofibrosis are fludarabine-busulfan (FB), originally reported by Slavin et al. [10], and fludarabine-melphalan (FM), originally reported by

G.S. and N.K. contributed equally as senior coauthors.
Financial disclosure: See Acknowledgments on page 6.

* Correspondence and reprint requests: Marie Robin, MD, PhD, Hematology Transplantation, Saint-Louis Hospital, University Paris 7, Paris, 1160 France.
E-mail address: marie.robin@paris7.jussieu.fr (M. Robin).

http://dx.doi.org/10.1016/j.bbmt.2016.02.019
1083-8791/

Giralt et al. [11]. Both regimens are reportedly as effective as and less toxic than myeloablative conditioning regimens. Previous studies have compared FB and FM conditioning regimens in various hematologic disorders. Shimoni et al. [12] reported that FM was more toxic and carried a higher NRM rate than FB, whereas more recently Baron et al. [13] concluded that FM and FB regimens led to similar rates of overall survival (OS) in patients with acute leukemia. Both studies concluded that compared with FB conditioning, FM conditioning induced better disease control, although this did not translate into improved disease-free survival owing to higher NRM. These studies did not include any patients with myelofibrosis, however, and there is scarce published evidence on the most effective RIC regimen in patients with myelofibrosis undergoing HSCT.

Owing to the median age of onset, the vast majority of patients with myelofibrosis will receive a RIC regimen. Two prospective studies in the setting of HSCT for myelofibrosis used either an FB or FM regimen. Kroger et al. [1] reported a 5-year disease-free survival rate of 51% in patients who received RIC with an FB regimen, whereas Rondelli et al. [14] reported that 35 of 66 patients remained alive following FM conditioning. Both studies found worse outcomes in patients undergoing unrelated donor transplantation. A recent Center for International Blood and Marrow Transplantation Research study that evaluated 5-year post-transplantation survival according to different conditioning regimens reported survival rates of 59% with FM, 46% with fludarabine-total body irradiation, 41% with FB, and 28% in other heterogeneous conditioning regimens; however, an advantage of the FM regimen was not confirmed in multivariate analysis, in which donor type remained the sole significant prognostic factor [6].

The aims of the present study were to identify whether significant differences exist between patients undergoing HSCT for myelofibrosis using either the FM or FB conditioning regimen, and whether the choice of regimen has an impact on transplant-related outcomes of progression-free survival (PFS), overall survival (OS), NRM, relapse rate, and incidence of graft-versus-host disease (GVHD). This study involved 2 large transplantation centers recruiting patients with myelofibrosis and using either the FM or FB conditioning protocol to facilitate a comparison of these regimens.

METHODS

Patients

A total of 160 consecutive patients age ≥18 years with primary or secondary myelofibrosis who received either an FB (Paris) or an FM (Hamburg) conditioning regimen between 2005 and 2014 were included in the study. Sixty-nine of these 160 patients have been reported previously [1,5]. Patients whose disease transformed to acute leukemia were excluded, as were patients enrolled in another ongoing prospective transplantation study. All patients provided consent for inclusion in this study in accordance with the Declaration of Helsinki.

Statistical Analysis

Outcomes in the 2 conditioning regimen groups were compared using Cox proportional hazards models (for OS and PFS) and proportional cause-specific hazards models (for NRM and relapse), both unadjusted and adjusted for patient age ≥55 years, donor type, female donor to male recipient, cytomegalovirus (CMV)-negative donor to positive recipient, CD34 cell dose, and intermediate-2 or high Dynamic International Prognostic Scoring System (DIPSS) score. A propensity score–based approach was used to limit confounding bias in the comparisons of outcomes in the 2 conditioning regimen groups. More specifically, inverse probability weighting (IPW) was applied to recreate 2 pseudo-populations in which the patients receiving FM or FB conditioning and antithymoglobulin-Fresenius would have similar characteristics [15,16]. For IPW analyses, each patient was thus weighted by the inverse of the estimated probability of receiving the conditioning that he or she actually received given his or her baseline characteristics. This probability, the so-called propensity score, was estimated using a logistic regression model with conditioning regimen as the dependent variable and patient age, time from diagnosis to transplantation, donor age, donor–recipient sex match, CMV match, CD34 cell enumeration in the graft, type of donor (HLA-matched sibling, unrelated, mismatched unrelated donor) and DIPSS score as predictors. Fractional polynomials were used to model the effects of patient and donor age and CD34 cell enumeration [17]. To investigate whether modern data mining methods would have led to a more accurate propensity score than logistic regression, boosted regression trees were also used, but these did not result in a better balance of weighted samples [18]. Standardized differences were used to compare the balance in baseline characteristics between groups, with a standardized difference <10% generally considered to indicate successful balance [19]. The overall balance in all baseline characteristics was measured using the C-statistic [20]. Models for weighted samples used a robust variance estimator [21,22]. Kaplan-Meier curves and weighted Kaplan-Meier curves were used to estimate the probability of OS and PFS, and cumulative incidence and weighted cumulative incidence curves were used to estimate the probability of NRM, relapse, and acute and chronic GVHD. Missing data were handled through multiple imputations by chained equations methods [23,24]. A total of 20 independent imputed datasets were generated and analyzed separately. Variables used for multiple imputation were not limited. Estimates of model parameters and weights obtained in the imputed datasets were then pooled over the imputations according to Rubin's rule [23].

RESULTS

Patient Characteristics

This 2-center study included a total of 160 patients, the characteristics of which are summarized in Table 1. The FM regimen consisted of fludarabine 90 mg/m^2 and melphalan 140 mg/m^2. The FB regimen consisted of i.v. busulfan (or oral equivalent) 8 mg/kg, fludarabine 180 mg/m^2, and antithymocyte globulin (ATG)-Fresenius 30 mg/kg for related donor graft recipients and 60 mg/kg for unrelated donor graft recipients. A total of 55 patients received FM conditioning (8 of whom also received in vivo T depletion with ATG-Thymoglobuline 15%), and 105 patients received FB conditioning and ATG-Fresenius. The most common pretransplantation treatments were hydroxyurea (in 49% of the FM group and 32% of the FB group) and ruxolitinib (in 9% of the FM group and 16% of the FB group).

JAK2 V617F status was available in 155 patients (97%), revealing mutations in 102 (67%). Among the 53 patients with nonmutated *JAK2* V617F, *MPL* status was determined in 43, 6 of whom demonstrated *MPL* mutations. Thirty-two of the 37 patients who were *JAK2* V617F and *MPL* negative were tested for calreticulin exon 9 (CALR) mutations, and 21 were positive; the other 11 patients were classified as "triple negative" (see additional table).

Spelenctomy was performed before HSCT in 53% of patients in the FM group, compared with only 4% in the FB group. The FM group was younger (median, 55 years versus 59 years), had a higher proportion of males (69% versus 55%), underwent HSCT after a longer duration of disease (median, 21 months versus 19 months), more often received a graft from a matched sibling donor (47% versus 18%), and received higher doses of CD34$^+$ cells (median, 7.5 × 10^6/kg versus 6.6 × 10^6/kg). GVHD prophylaxis was based on calcineurin inhibitors, mainly cyclosporin, combined with either methotrexate or mycophenolate mofetil in both groups.

The 2 groups displayed marked dissimilarities in many variables, as exemplified by the standardized differences (Table 1); however, IPW was successful in reducing the imbalance for most baseline characteristics even if a perfect balance was not achieved, and the overall imbalance in propensity score was much lower in the weighted samples than in the original sample (C-statistic, 0.5 of 0.083 versus 0.375). Of note, residual imbalances remained for the time