

Impact of comorbidity indexes on non-relapse mortality

Aliénor Xhaard¹, Raphael Porcher²⁻³, Jason W. Chien⁵,
Régis Peffault de Latour¹⁻⁴, Marie Robin¹, Patricia Ribaud¹,
Vanderson Rocha¹, Agnès Devergie¹, Christelle Ferry¹,
Paul J. Martin⁵ and Gérard Socié¹⁻⁴

¹ AP-HP, Hôpital Saint-Louis, Service d'Hématologie-Greffe,
Paris, F-75010, France

²Université Paris VII, IUH-IFR105, Paris, F-75010, France,

³Inserm, U717, Paris, F-75010, France,

⁴Inserm, U728, Paris, F-75010, France,

⁵Clinical Research Division, Fred Hutchinson Cancer Research Center,
Seattle, WA, USA

Word count: Abstract: 200 ; Text words: 3295.

**Key words: comorbidities, allogeneic hematopoietic stem cell
transplantation, non-relapse mortality, overall survival**

Running title: Comorbidities & NRM after allogeneic SCT

Correspondence to: Professeur Gérard Socié, MD, PhD

Service d'Hématologie Greffe, Hôpital Saint Louis

1 Av Vellefaux, 75010 Paris, France.

Phone: +33 1 42 49 98 24

Fax: +33 1 42 49 96 36

E-mail: gerard.socie@sls.aphp.fr

Abstract

Comorbidity indexes (CI) have been reported to predict non-relapse mortality (NRM) and overall survival after allogeneic hematopoietic stem cell transplantation (HSCT) [Charlson's Comorbidity Index (CCI), Hematopoietic Cell Transplantation CI (HCT-CI) & the Pre-transplantation Assessment of Mortality (PAM) score]. Which of these indexes best predict survival, is unknown, yet.

We retrospectively studied 286 patients who underwent allogeneic HSCT. HCT-CI and PAM required grading according to pre-transplant pulmonary function tests (PFTs), which were lacking for some patients. We thus designed a reduced HCT-CI and an adjusted PAM, without results of PFTs.

Using CCI, 25% of patients had indexes of 1 or more; median reduced HCT-CI was 1; median adjusted PAM score was 24. The discriminative properties of the 3 CI were rather low in our population. Comparison of patients- and transplant-characteristics between our and Seattle group's cohorts however revealed significant differences with more children, more cord blood HSCT and HSCT for Fanconi Anemia in Saint Louis. Finally, multivariate analysis of scoring items revealed that age, matched unrelated or mismatched donor and hepatic disease were associated with NRM in our cohort.

Translating use for patients counseling or decision to proceed to transplant of these CI will need prospective studies in a large independent cohort.

Introduction

Allogeneic hematopoietic stem cell transplantation (HSCT) is a potentially curative procedure for many patients with hematological malignancies and some non-malignant hematological diseases. The development of reduced-intensity conditioning regimens, the improvement of supportive care and modifications introduced in immunosuppressive therapies allowed to extend the indications of allogeneic HSCT to older, less fit and more heavily pretreated patients. However, this treatment procedure remains associated with a significant mortality rate, either due to relapse or to non-relapse causes. Non-relapse mortality (NRM) is mainly associated with acute and chronic graft-versus-host disease (GVHD), infections, interstitial pneumonia¹ and toxicities of the transplantation procedure. The estimation of overall survival (OS) and of NRM is of primary importance in weighing the risk/benefit ratio of allogeneic HSCT and for patient counseling.

However, standard factors such as age, type and stage of the underlying hematological disease and characteristics of the transplantation procedure lack discriminatory power and are insufficient to predict transplantation tolerance² and post-transplantation survival. Moreover, numerous studies have shown that comorbidities affect patients' outcome, especially in cancer patients. Comorbidities have an impact on both the risks of comorbid and global death^{3,4}. While some transplantation centers evaluate cardiac and pulmonary functions before allogeneic HSCT, there have been very few studies proving the usefulness of this pre-transplant cardiopulmonary evaluation^{5,6}. Moreover, no single-organ comorbidity used alone has a sufficient prognostic value to predict outcome after allogeneic

HSCT⁶. Consequently, comorbidity scales have been developed, combining various types of comorbidities known to individually influence outcome. As not all comorbid conditions have the same impact on patient's outcome, they are differently weighted according to their severity and to their impact on patient's outcome⁷.

The most studied comorbidity scale is the Charlson Comorbidity index^{3,8} (CCI). The CCI has been widely used to predict mortality in a variety of medical conditions, including solid tumors and hematological malignancies⁴. In the specific context of allogeneic HSCT, the CCI was demonstrated to be useful to predict NRM and OS among allogeneic HSCT recipients^{9,10}. The predictive value of the CCI was better for patients aged less than 40 years¹¹. However, it was reported that the CCI lacked sensitivity in the context of allogeneic HSCT since comorbidities were found in only 13% of the studied patients¹². This lack of sensitivity was found for related and unrelated donors, whichever the conditioning regimen was (myeloablative or non-myeloablative)¹². Subsequently, the CCI was modified to develop a new comorbidity scale specifically designed to capture chronic medical conditions and comorbidities frequently encountered in allogeneic HSCT patients and implicated in the evaluation of non-relapse mortality. This new scoring system, the hematopoietic cell transplantation comorbidity index (HCT-CI), was the first comorbidity index specifically designed for allogeneic HSCT¹². The HCT-CI appeared to be more sensitive than the CCI in capturing patients with comorbidities, in distributing them into risk groups and in predicting non-relapse mortality and overall survival during the first 2 years after allogeneic HSCT¹².

In the context of allogeneic HSCT, another score was developed, designed at predicting all-cause mortality during the first 2 years after allogeneic HSCT. This pre-transplantation assessment of mortality (PAM) score takes into account informations about the underlying hematological disease, the transplantation procedure and the age of the patient in association with some comorbidities¹³. This score was designed with only 8 variables, making it simple and easy to use. Higher PAM scores are associated with decreased 2-year overall-survival.

In both studies, patients from the same institution were used to develop and validate the scores. Although there were some differences in the training and the validation cohorts and sometimes a decline in the discriminatory ability for NRM (for the HCT-CI) between the two samples, the indexes continued to strongly predict outcome. However, the HCT-CI and the PAM score need to be tested and validated on an external population. In the present study, we tested the CCI, the HCT-CI and the PAM score on patients who underwent allogeneic HSCT in our centre.

Design and methods

Patients

Patients undergoing a first allogeneic HSCT at Saint-Louis hospital, France, between April 2004 and December 2006 were considered. A total of 286 patients were included in the study. Patients-, disease- and transplant characteristics are detailed in **Table 1**. Median follow-up of surviving patients is 18 months (min: 18 days, max: 37 months). Diseases were classified into risk categories according to the HCT-CI and the PAM score with slight modifications (congenital dyserythropoietic anemia, sickle cell

disease and thalassemic syndromes were classified as low-risk diseases, and Fanconi anemia, paroxysmal nocturnal hemoglobinuria and immunodeficiency syndromes were classified as high-risk diseases). Matching between donors and recipients was determined according to HLA-A, -B, -C, -DQ and -DR compatibility. HLA-A, -B and -C were determined by either intermediate resolution DNA typing or high-resolution techniques. HLA matching for -DRB1 and -DQB1 was done at the allelic level.

Data collection

Data needed for the classification according to the CCI, the HCT-CI and the PAM score were extracted from detailed review of the patients' medical reports. Laboratory values were collected before the beginning of the conditioning regimen. Pulmonary function tests data (PFTs) and evaluation of left ventricular ejection fraction (LVEF) were performed in the month preceding allogeneic HSCT. As follow-up was shorter than 5 years, patients were given scores based on the original CCI, not including age, as suggested by Charlson et al⁹. Comorbidities of the CCI, the HCT-CI and the PAM score were graded as published in the original description reports^{3,8,12,13}. The two indexes specifically developed for HSCT required grading according to pulmonary function exploration; since this was missing in a significant number of our patients, a reduced version of the HCT-CI (without PFTs) and an adjusted version of the PAM (multiplication by 1.25 if grading on a 40-scale instead of a 50-scale) were computed.

Statistical analysis

Differences in characteristics between samples of original reports and the present sample were assessed in terms of standardized difference, which

is the absolute mean difference divided by the pooled standard deviation, expressed as a percentage (shown in **Table 1**).

Survival curves were estimated using Kaplan-Meier estimator, and cumulative incidence curves of NRM according to published methods, treating relapse as competing event¹⁴. The likelihood ratio statistics in Cox proportional hazards models was computed for each index, with its associated P-value, as a measure of association between the comorbidity indexes collapsed into risk groups and the outcomes. Discriminative performance of comorbidity indexes was then evaluated by the c-index¹⁵. Standard errors of c-indexes were computed from that of Somers' D_{xy} rank correlation coefficient¹⁶.

Additionally, Cox proportional hazards models were fitted to estimate in our sample the association with the outcomes of each single factor part of a comorbidity index, both in univariate and multivariate analyses. Factors with less than five observations in the sample were ignored in this analysis, as well as factors for which no event occurred, to avoid problems with model fit. Results were expressed in terms of either parameter estimate and standard error (β [se]) or hazard ratio (HR), to facilitate comparison with original reports.

Analyses were performed using R statistical software (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Grading according to the 3 indexes

The data needed for the grading according to the CCI were available for all the 286 patients. 75% of our patients were in the low risk category (0) whereas 19% were in the intermediate (1) and 6% in the high (≥ 2) risk group category (**table 2**). Patients in the high risk group (CCI score ≥ 2) had worse 2-year OS and higher 2-year NRM than patients with scores of 0 or 1 (**figure 1A-B**). However, the CCI was not discriminative for OS or NRM in patients having a score < 2 .

We could apply the original version of the HCT-CI to 86 patients only, because of missing PFTs data. The distribution in risk groups of the 86 patients graded in the HCT-CI was as follow: low (0): 6%, intermediate (1-2): 15%, high (≥ 3): 79% (among them, 34% had score of 3 and 45% had score ≥ 4) (**table 2**). We applied a reduced HCT-CI scale (excluding pulmonary comorbidities) to the remaining 182 patients (18 missing data of heart valve disease, in patients for whom LVEF was evaluated using cardiac scintigraphy and not echography) and pooled these data with the original HCT-CI of 86 patients above mentioned. 30% of the patients were in the low risk group (reduced HCT-CI=0), whereas 47% were in the intermediate (1-2) and 23% in the high-risk group (≥ 3) (**table 2**). The reduced HCT-CI did not appear discriminative in our cohort (**figure 1C-D**): neither the 2-year OS, nor the 2-year cumulative incidences of NRM were different in the 3 risk groups individualized following the reduced HCT-CI. There was no association of higher reduced HCT-CI scores with worse OS or increased NRM. However, we confirmed the higher sensitivity of the HCT-CI to detect comorbidities compared to the CCI: whereas 75% of our patients had a score of 0 in the

CCI, they were only 6% with the original version of the HCT-CI and 30% in the reduced version.

The original version of the PAM score was applicable to only 81 patients and the adjusted version (multiplication by 1.25 after a 40 points grading, this latter one excluding FEV1 and DLCO) was applied to the remaining 205 patients. Distribution into risk groups according to the PAM and to the adjusted PAM is presented in **table 2**. Stratification into risk groups according to the adjusted PAM allowed us to identify 3 groups with distinct 2-year OS (**figure 1 E-F**). However, the adjusted PAM did not predict NRM in our cohort.

As, until recently, PFTs were not performed for our patients before allogeneic HSCT per protocol but according to the physician's suspicion of altered lung function, we performed another analysis assuming that PFTs were not assessed before allogeneic HSCT because of low probability of altered lung function, and we assigned patients with missing values to the best lung function category of the HCT-CI and the PAM. This did not modify significantly the results and did not lead to improve the discriminative properties of the HCT-CI and of the PAM in our patient cohort.

Statistical comparison of the 3 indexes

The discriminative properties of the 3 indexes were rather low in our population. C-indexes for OS were of 0.514, 0.499 and 0.579 for the CCI, reduced HCT-CI and adjusted PAM score, respectively. Based on the likelihood ratio, there was no association of the reduced HCT-CI with OS and NRM in our cohort. Compared to the reduced HCT-CI, the likelihood ratio and the c-index were higher for the CCI, for both OS and NRM. Based on the likelihood ratio, there was an association of higher adjusted PAM scores with

worse OS. For OS and NRM, the c-index was significantly higher for the adjusted PAM score than for the CCI and the reduced HTC-CI.

Grading of our patients according to the 3 different scales showed that the CCI identified a high-risk group, with decreased OS and increased NRM. However, the CCI was of limited value because this high-risk group corresponded to less than 10% of our cohort. The HCT-CI, either in its original version or as reduced version excluding the pulmonary comorbidities, was not significant in our cohort. The adjusted PAM score identified three distinct risk groups for OS.

Study of non relapse mortality

Finally since none of the previously described indexes reliably predict NRM in our patient's cohort, we ran multivariate analyses taking into account single factors and potent detrimental effects of each parameter of the different indexes. Results are displayed in **tables 3 and 4**. Aside from classical factors that include older age and transplant from matched unrelated or mismatched donor, only hepatic disease as defined in the CCI proved to predict NRM in our patient's cohort, by multivariate analysis. Excluding patients with Fanconi anemia, thalassemic syndromes and sickle cell disease which are over-represented in the Saint Louis cohort did not affect the results significantly.

Discussion

In this study we aimed to test and validate the HCT-CI and the PAM score on an external population, and to know which of these two indexes best predict HSCT outcomes in our patient's population. Unfortunately the HCT-CI did not predict OS or NRM in our population. Although the PAM was

moderately predictive of OS, it did not perform as well as it did in the original study. Furthermore, the PAM was not designed to predict NRM, and therefore, could not be assessed for NRM in our population. Indeed only liver disease as defined by the CCI add in predicting NRM better than the classical age and donor type factors.

As pulmonary comorbidities are heavily weighted in the HCT-CI and have been demonstrated to significantly influence post-transplant outcome, it is not surprising that the reduced HCT-CI (which excluded PFTs data) is not predictive in our cohort. However, the adjusted version of the PAM score still discriminates 3 risk groups for OS. We hypothesized that the lack of pulmonary comorbidities was not the only reason for which the HCT-CI is not discriminative in our cohort. We then searched for reasons to explain these differences with the original studies of the HCT-CI and the PAM.

We compared patients' characteristics in the different studies and this revealed significant differences. In our cohort, we included children below 15 years of age, representing 20% of our population. Children were also included in Sorrow's cohort (unknown percentage), whereas there was no patient below 15 years of age in Parimon's study. Concerning the underlying hematological disease, our cohort included less CML (6% compared to 20% by Sorrow and 36% by Parimon) and less MDS (5% compared to 19% by Sorrow and 15% by Parimon) but more ALL, more non-malignant disease and some Fanconi anemia. Our cohort also included 14% of cord blood transplant (7% of single- and 7% of double-cord blood transplant), whereas there was not any one in Sorrow's cohort and only 1% in Parimon's development cohort (data non available for the validation cohort). Peripheral blood stem cells were used as hematopoietic stem cell source for 71% of the

patients in Sorrow's study, whereas it represented only 23% of the development cohort of Parimon and 43% of our population. Patients were allografted from 1997 to 2003 in Sorrow's study and from 1990 to 2002 in Parimon's one, and we recorded data of patients allografted from April 2004 to December 2006. This might contribute to the apparent difference in mortality rates between the different cohorts. Our 2-year mortality rate of 34% is lower than Parimon's one (51% for the entire validation cohort) and than Sorrow's one, which might have influenced the performance of the predictor model. It was also worth investigating that the association of donor type and outcomes may be confounded by disease risk. However, both variables appeared in the multiple model developed in the adults with haematological malignancies, so the possible confounding was removed by adjustment. Nevertheless, we tested for a possible interaction between both, that was not significant ($p=0.56$). In the whole cohort, disease risk was not kept by the model selection procedure.

Distribution of the comorbidity indexes according to the type of conditioning regimen showed that there were more non myeloablative than myeloablative patients which had a CCI ≥ 2 (11% of non myeloablative versus 3% of myeloablative). In the original study of the HCT-CI, non myeloablative patients had higher comorbidity scores¹². In our cohort, for the original version of the HCT-CI, there were no differences in the repartition of the patients in the different risk groups. More ablative than non myeloablative patients had PAM scores (either as original or adjusted version) corresponding to the low risk category.

The frequency of comorbidities also slightly differs between populations. In Sorrow's study, the most prevalent comorbidities are mild

(16%) and moderate (24%) pulmonary abnormalities, mild hepatic abnormalities (16%), mild renal abnormalities (10%) and hypertension (10%). In our population, the most frequently encountered comorbidities are moderate and severe pulmonary abnormalities (28% and 47%, in the population of patients having performed PFTs before the allogeneic HSCT), all hepatic abnormalities (mild: 32%, moderate/severe: 16%), and also infection (which was present in 23% of our patients versus only 4% of patients in Sorror's study).

To demonstrate its ability to predict outcome in other patients' groups, the HCT-CI was tested in a cohort of patients with AML in first complete remission undergoing allogeneic HSCT and transplanted either at the MDACC or at the FHCRC¹⁷. In this study, the HCT-CI was demonstrated to be more sensitive than the CCI to capture comorbidities and to predict outcome. However, multivariate analysis of risk factors for OS and NRM showed that there were no differences between the intermediate and the high-risk HCT-CI groups for both outcomes in the MDACC cohort, whereas the 3 HCT-CI risk groups of the FHCRC cohort presented significant differences for OS and NRM. Moreover, this same multivariate analysis didn't show in the MDACC cohort any influence on outcome of classical factors such as age, donor type and cytogenetics.

We studied the prognostic value of each comorbidity factor included in the CCI and in the HCT-CI. In the CCI, the comorbidities whose prevalence was high enough to study the predictive value were: hypertension, congestive heart failure, cerebrovascular disease, mild hepatic disease and solid tumor

without metastases. Univariate analysis showed that none of these variables were significantly predictive for OS or NRM. This is quite surprising because the CCI allowed us to discriminate one high-risk group with altered OS and higher NRM.

Concerning the HCT-CI, we studied the prognostic value of comorbidities whose prevalence was higher than 3% (cardiac disease, cerebrovascular disease, psychiatric disturbance, hepatic disease, pulmonary disease, infection and heart valve disease). In the HCT-CI design study, only comorbidities with an adjusted multivariate hazard ratio (HR) > 1.3 were taken in account. Applying this to our population, psychiatric disturbance was significant in multivariate analysis for OS and NRM (with HR of 1.5 and 1.6 respectively), and heart valve disease was a significant predictor for OS (HR of 1.5 by multivariate analysis). The other comorbidities were not predictive for OS or NRM, even the newly included item “infection”.

For the PAM score, we compared the data of the original PAM study with ours. In our cohort, 40% of patients received a non myeloablative conditioning regimen (versus 6% in the PAM study) and none of our patients received more than 12 Gy TBI (compared to 35% in the PAM study). We compared the scores attributed by Parimon to the different comorbidity items of the PAM score to the scores we calculated for the very same items in our population. The number of patients was insufficient to conclude for the items: age above 60, TBI>12Gy and serum creatinine level > 1.2mg/L. Results for the other items are reported in **table 5**.

One major limitation of our study is the absence of pre-transplant pulmonary function tests (PFTs) in 70% of the study population. Until

recently, pre-transplant PFTs were not systematically performed in all our patients although post-transplant PFTs were systematically performed around day 100 post-transplant to detect early alterations of PFTs post transplant. The unavailability of baseline PFTs in a high percentage of our population did not allow us to calculate properly the HCT-CI and the PAM score in our population. The adjusted version of the PAM was, nevertheless, a significant predictor of overall survival in our population. Another limitation of our study is that the data were collected retrospectively on the basis of the review of the patients' medical reports.

Conclusion

In conclusion, this study confirms that the HCT-CI detects comorbidities in a higher percentage of patients than the CCI does. In our cohort, the CCI identified a small high-risk subpopulation with decreased OS and increased NRM after allogeneic HSCT. A reduced version of the HCT-CI was not discriminative either for OS or for NRM. An adjusted version of the PAM excluding FEV1 and DLCO allowed to discriminate 3 risk groups with distinct 2-year overall survival. The HCT-CI has been endorsed by the CIBMTR and data will be prospectively collected on thousands of patients, without previous prospective validation. Thus, a study aiming to prospectively study the impact of comorbidities in several centers in Europe and in the USA is timely and warranted for patient counseling and comparisons of results from different groups.

References

1. Bacigalupo A, Sormani MP, Lamparelli T, Gualandi F, Occhini D, Bregante S et al. Reducing transplant-related mortality after allogeneic hematopoietic stem cell transplantation. *Haematologica* 2004; **89**; 1238-47.
2. Artz AS, Pollyea DA, Kocherginsky M, Stock W, Rich E, Odenike O et al. Performance status and comorbidity predict transplant-related mortality after allogeneic hematopoietic cell transplantation. *Biol Blood Marrow Transplant* 2006; **12**; 954-64.
3. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**; 373-83.
4. Extermann M. Measurement and impact of comorbidity in older cancer patients. *Crit Rev Oncol Hemato* 2000; **35**; 181-200.
5. Bolwell BJ. Are predictive factors clinically useful in bone marrow transplantation? *Bone Marrow Transplant* 2003; **32**; 853-61.
6. Alamo J, Shahjahan M, Lazarus HM, de Lima M, Giralt SA. Comorbidity indices in hematopoietic stem cell transplantation: a new report card. *Bone Marrow Transplant* 2005; **36**; 475-9
7. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol* 2003; **56**; 221-9

8. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994; **47**; 1245-51.
9. Diaconescu R, Flowers CR, Storer B, Sorrow ML, Maris MB, Maloney DG, et al. Morbidity and mortality with nonmyeloablative compared with myeloablative conditioning before hematopoietic cell transplantation from HLA-matched related donors. *Blood* 2004; **104**; 1550-8.
10. Sorrow ML, Maris MB, Storer B, Sandmaier BM, Diaconescu R, Flowers C, et al. Comparing morbidity and mortality of HLA-matched unrelated donor hematopoietic cell transplantation after nonmyeloablative and myeloablative conditioning: influence of pretransplantation comorbidities. *Blood* 2004; **104**; 961-8.
11. Shahjahan M, Alamo J, de Lima M, Khouri I, Gajewski J, Andersson B et al. Effect of comorbidities on allogeneic hematopoietic stem cell transplant outcomes in AML/MDS patients in first complete remission. *Biol Blood Marrow Transplant* 2004; **10 Suppl 1**; S12-13.
12. Sorrow ML, Maris MB, Storb R, Baron F, Sandmaier BM, Maloney DG et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood* 2005; **106**; 2912-9.
13. Parimon T, Au DH, Martin PJ, Chien JW. A risk score for mortality after allogeneic hematopoietic cell transplantation. *Ann Intern Med* 2006; **144**; 407-14.

14. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Wiley & Sons, editor. New York (NY); 1980: 163-178.
15. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984; **3**: 143-52.
16. Harrell FE, Jr. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. Springer-Verlag, editor. New York (NY); 2001: 249-253
17. Sorrow ML, Giralt S, Sandmaier BM, De Lima M, Shahjahan M, Maloney DG et al. Hematopoietic cell transplantation specific comorbidity index as an outcome predictor for patients with acute myeloid leukemia in first remission: combined FHCRC and MDACC experiences. *Blood* 2007;**10**: 4606-13.

Legend to Figure

Figure 1: overall survival and non-relapse mortality in the different risk-groups defined by the CCI (A-B), the reduced HCT-CI (C-D) and the adjusted PAM (E-F)

