Supplementary Data:

Variables used in the multivariate models

In this work, we have applied a similar methodology than the one used by CIBMTR for the annual report of survival outcome for all centres in their network (http://www.cibmtr.org). In contrast to the CIBMTR, the covariables that were considered here did not include ethnic group (not very relevant for Belgium) nor comorbidity scores (almost impossible to retrieve retrospectively).

Disease Risk Index

The Disease Risk Index (DRI) was adapted from Armand et al[21]. It relies on 3 parameters:

- The Disease (D) parameter is defined according to the WHO classification of the hematological malignancies;
- The Stage (S) parameter is based on a EBMT ProMISe (Project Manager Internet Server) variable explaining the status of the disease in combination with the different D groups.
- The Genetics (G) parameter defines main genetic risk groups only for the D-groups of AML and MDS. This variable was part of the Med-B form and therefore not mandatory. As a result it was hardly filled out in the ProMISe database. For this project, the BCR has therefore actively contacted the transplant centres to retrieve genetic information.

The resulting DRI variable has four different levels: Low, Int(ermediate), High and Very High for different combinations of D, S and G.

Disease (D)	Genetics (G)	Status (S)	DRI Group
AML	Favorable	CR	Low
		Advanced	High
	Intermediate	CR	Int
		Advanced	High
	Adverse	CR	High
		Advanced	Very high
ALL		CR1	Int
		CR2	High
		CR3	High
		Advanced	Very high
MDS Low-risk	Intermediate	Early	Int
		Advanced	Int
	Adverse	Early	Int
		Advanced	High
MDS high-risk	Intermediate	Early	Int
		Advanced	High
	Adverse	Early	High
		Advanced	High
Myeloproliferative neoplasm		Any	Int
CML		Chronic phase 1/2	Low
		Accelerated phase	Int
		Blast phase	Very high
CLL		CR	Low
		PR	Low
		Advanced	Int
Indolent NHL		CR	Low
		PR	Low
		Advanced	Int
Mantle cell lymphoma		CR	Low
		PR	Int
		Advanced	High
T-cell NHL		CR	Int
		PR	Int
		Advanced	High
Aggressive NHL		CR	Int
		PR	Int
		Advanced	Very high
Burkitt lymphoma		CR	High
		PR-Advanced	Very high
Hodgkin lymphoma		CR	Low
		PR	Int
		Advanced	High
Multiple myeloma		CR/VGPR/PR	Int
		Advanced	High

CR: complete remission; PR: partial remission

Genetic risk groups

The genetic risk groups include cytogenetics and molecular anomalies. They are only defined for patients with the diagnosis of acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS).

They were adapted from ELN classification for AML[22,23] and from IPSS score for MDS[24].

Diagnosis	Group	Abnormality
AML (adapted from ELN classification)[20,21],	Favorable	Isolated CBF: t(8;21) / *RUNX1-RUNX1T1* inv 16 / t(16;16) / *CBFB-MYH11* t(15;17) / *PML-RARA* Normal karyotype (N) & mutated *NPM1* without *FLT3*-ITD N karyotype & mutated *CEBPA* Normal (N) karyotype & mutated *NPM1* without *FTL3*-ITD
	Intermediate	Any other abnormalities (abn)
	Adverse	Complex (≥3 abn) without CBF Monosomal karyotype Rearranged *KMT2A* / t(v;11) except t(9;11) / *MLLT3-KMT2A* inv(3) / t(3;3) / *RPN1-EVI1* or mutated *EVI1* Abn (17p); -7;-5 or del(5q) t(6;9) / *DEK-NUP214* t(9;22) / *BCR-ABL1*
MDS (adapted from the IPSS score)[22]	Standard (IPSS low or intermediate)	N -Y del(5q) del(20q) +8 Any other abn
	Adverse (IPSS high)	Complex (≥ 3 abn) Abnormal 7

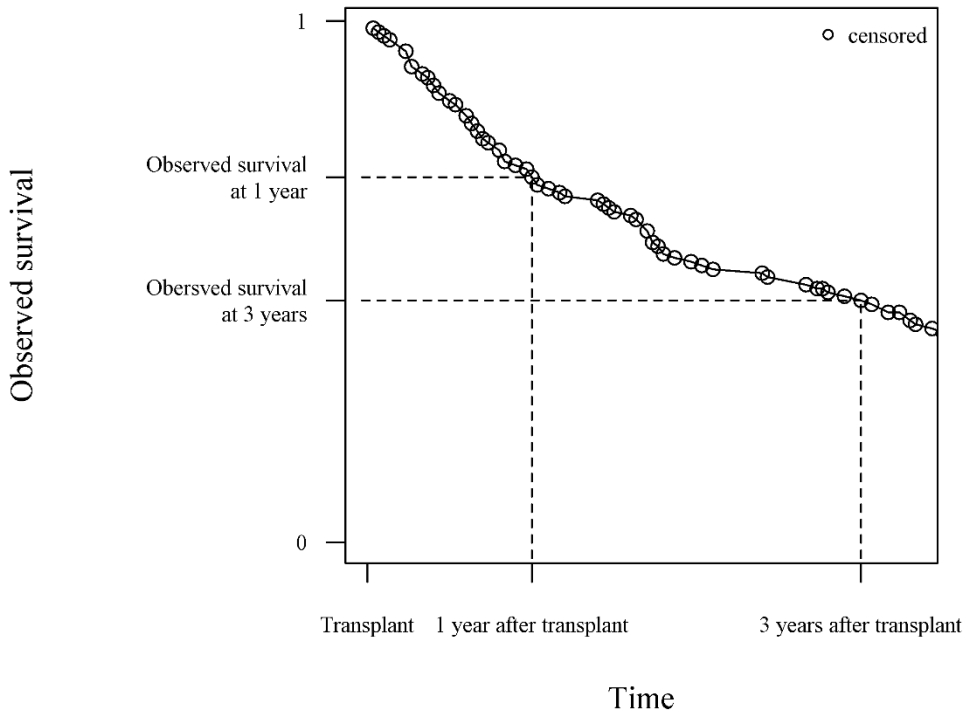
CBF: Core binding factor; RUNX1: Runt-related transcription factor-1; RUNX1T1: Runt-related transcription factor-1 translocation partner 1; PML: Promyelocytic leukemia; RARA: Retinoic acid receptor alpha; CBFB: core binding factor subunit B; MYH11: Myosin-11; NPM1: Nucleophosim; FTL3: FMS-like tyrosine kinase 3; ITD: Internal tandem duplication; CEBPA: CCAAT/enhancer binding protein alpha; BCR: Breakpoint cluster region; ABL: Abelson murine leukaemia viral oncogene homolog 1; NUP: Nucleoporins; MLL: Mixed lineage leukaemia; EVI: Ectotropic virus integration site 1; RPN: Ribophorin

Why EBMT risk score is not included in the model

Since the EBMT risk score is created by directly combining variables we wish to account for, a statistical multivariable model containing the EBMT risk score should not contain the variables on

which the EBMT risk score is based in order to avoid multicollinearity. For every data selection and modelling exercise we have performed a model with EBMT risk score according to Grathwohl A [The EBMT risk score. Bone Marrow Transplantation 2012 47: 749–756] (and therefore without Armand's DRI, age, time since diagnosis) and one without EBMT score (consequently with all other variables). The final results of models with and without EBMT risk score did not differ much. Given this finding and since models with EBMT risk score could show convergence problems, it was decided to present only models without the EBMT risk score.

Details of the Generalized linear regression model (GRp model) on pseudo values



Schematic graph of a survival curve : the Cox model focuses on the whole curve while the generalized logistic model focuses on one time point, 1-yr or 3-yr in our study.

Firstly, we have built 2 models for standardisation. 1 model to predict observed survival at 1 year after transplant time and another to predict observed survival at 3 years after transplant time.

These models were selected by using a backward selection based on the corrected Akaike criterion (AICc)[27]. The goal of the corrected Akaike criterion is to obtain a model which fits at best the data (ie, shorter distance between the predicted and the observed values) with a minimum of variable. The formula of AICc is the following:

$$AIC_c = -2\ln(l) + 2K + \frac{2K(K+1)}{n-K-1}$$

Where n = Number of observations, K = Number of model parameters and l = likelihood of the model

The more the predicted values fits to the observed values, the smaller $-2\ln(l)$ is small and the more parameters there are, the larger $2K + \frac{2K(K+1)}{n-K-1}$ is. The AICc criterion has to be minimized to select the best model. (NB : the correction of the Akaike criterion has been made for small samples. However the corrected and non-corrected values do not really differ on our dataset due to the high number of transplants.)

Secondly, the 3 centre related-variables were added to the models which has been selected for standardisation, in order to evaluate their link with the observed survival at 1 year and at 3 years.

We applied a fixed effects censored data generalized linear regression model (GRp model)[25] to fairly compare survival estimates across the different Belgian transplant centres after adjustment of the overall survival rates for the main prognostic covariables (i.e. the case-mix).

By using the pseudo-observations, you create so called ‘probabilities’ for every patient and these can indeed be modelled by a more standard regression model, using e.g. a logit link. The effect of covariates on the survival at a fixed time point is studied as such. Note that by doing so and by going through the pseudo-observations via the Kaplan-Meier estimator you smartly combine time

to event and censoring into one single observation (the pseudo-observation) per patient. So all information is kind of kept, yet modelling can be performed.

The traditional cox model, on the other hand, links the covariates to the entire hazard function h(t) and provides a way to go from hazard to survival function S(t). So in that case the effect on the entire survival course is modelled and not only at a certain time point (e.g. 1-year).

Definition of the pseudo-observations

- Compute firstly the Kaplan-Meier estimate of survival at fixed time t (1 or 3- years in our case) based on the entire sample: $\hat{S}_p(t)$

- Next, compute the Kaplan-Meier estimate of survival at t based on the entire sample minus observation i: $\hat{S}_p^i(t)$

- For every recipient i, a pseudo value can now be defined:

$$\hat{\theta}_i = n\hat{S}_p(t) - (n-1)\,\hat{S}_p^i(t)$$

Without the presence of censoring before 1- or 3-years, the pseudo-observations would simply become the indicators stating the individuals were alive or dead at 1- or 3-years. In this work and thus in the presence of censoring (before 1- or 3-years), all individuals with a time to event beyond the time of interest (1- or 3-years resp.) have the same pseudo-observation, regardless of whether they die, survive or are lost to follow-up beyond the time of interest. This particular pseudo-observation would be the maximum of pseudo-observations in the data set. This value is typically close to 1. The individuals experiencing their event (death or censoring) before the time of interest all have different pseudo-observations (changing as defined by the formula above) that are all smaller than the maximum pseudo-observation of the previous set of individuals experiencing their event beyond the time of interest. The pseudo-observations are not constrained between the interval [0,1] but are typically close to 0 and 1 even if they can be outside the range.

Model building

Since the pseudo-observations are not purely binomial (taking only values of 0/1), a regular logistic regression assuming a binomial distribution is not applicable to model the relation between them and the list of predictors (covariables) we are interested in. According to Logan et al.[28], the logit link function can be applied however. In this case, it will model the mean of the pseudo-observations and will not be applied on the pseudo-observations directly. So, the GRp model relates the pseudo-observations to a list of covariables by means of a generalized linear regression model with the following link function:

$$g(\theta_i) = log\frac{\theta_i}{1-\theta_i} = \beta_0 + \sum_{l=1}^{p}\beta_l Z_{il}$$

All covariables of interest were entered in the GRp model (resp. for 1-year and 3- year Survival probabilities) and a backward elimination procedure based on Akaike's information criterion (AICc) was used to define the list of covariables that predicted the survival (at resp. 1-year or 3-years) the most strongly and that thus had to be retained in the final model. The results of each of the AICc backward selection procedures A final model with the p retained covariables was fit (no centre effect). The model now estimates the likelihood of survival (or inversely death) in the first year (resp. 3 –years) for every patient.

Centre Case Mix Scores

Depending on his or her particular values for the covariables, a patient may be estimated by the GRp model to have a high or low likelihood of surviving the first year (resp. 3rd year). The linear combination of the covariable values as given by the GRp model ($\beta_0 + \sum_{l=1}^{p}\beta_l Z_{il}$) can thus be seen as a score: high implying a high likelihood of survival in the first year (resp. 3rd year), low implying a low likelihood of survival in the first year. For each centre, a centre case-mix score can now be computed by averaging the risk scores for all recipients recorded for a centre:

$$CMS_j = \frac{1}{n_j}\sum_{i \in C_j}\left(\beta_0 + \sum_{l=1}^{p}\beta_l Z_{il}\right)$$

Note that n_j stands for the number of recipients at centre j and C_j is the set of recipients recorded for centre j.

Centre specific predicted Survival

Given a recipients' characteristics, a score can thus be calculated from the fitted GRp model which in its turn and by the GRp model leads to an estimated survival rate for the recipient:

$$\hat{p_i} = \frac{\exp(\beta_0 + \sum_{l=1}^{p}\beta_l Z_{il})}{1 + \exp(\beta_0 + \sum_{l=1}^{p}\beta_l Z_{il})}$$

In order to obtain the predicted survival rate at a centre j, the average of all estimated survival rates of its recipients C_j is taken:

$$E(S_j) = \frac{1}{n_j}\sum_{i \in C_j}\hat{p_i}$$

Confidence Limits

Confidence limits of the Centre specific predicted Survival estimate are calculated by a bootstrapping methodology as described in Logan et al. [24]. In short it is based on the Pearson residuals obtained for every recipient i by $r_i = \frac{\hat{\theta}_l - \hat{p}_l}{\sqrt{\hat{p}_l(1-\hat{p}_l)}}$. A bootstrapping algorithm is then carried out on the residuals as follows:

- For b = 1 to 10.000, generate r_i^{*b} for recipient i by sampling with replacement from the set of Pearson residuals of all recipients contributing to the model

- For every of the 10.000 runs, for a single recipient i, calculate bootstrap estimated survival rates:

$$Y_i^{*b} = \hat{p}_l + r_i^{*b}\sqrt{\hat{p}_l(1-\hat{p}_l)}$$

- Now compute 10.000 centre specific predicted survival rates for each centre j:

$$j: S_j^{*b} = \frac{1}{n_j} \sum_{i \in C_j} Y_i^{*b}$$

As a consequence every centre j has 10.000 bootstrap estimates of the centre specific predicted survival rates. These are used to construct 95% predicted confidence bounds for centre 's specific predicted survival by taking the 2.5th and 97.5th percentile of all S_j^{*b} across the 10.000 bootstrap samples.

Since the centre itself is not entered as a covariable in the GRp model, centre j's specific estimated survival rate represents the survival rate that might be observed at centre j if there were no centre effects and as if the recipients of centre j have been transplanted at any centre in the list of centres that are compared (i.e. the expected survival[28]). The 95% intervals around this value give a range of values that is likely to include the expected survival (adjusted for the case-mix). The observed survival (given here by the unadjusted Kaplan-Meier estimates) can be compared to get an impression of how far off they are of the expected results. We state that when the observed survival lies outside of the bootstrap-predicted 95% intervals, the centre is over- or under-performing the overall network of centres, according to[28]

Odds ratio

Based on the methodology of the GRp model[25], the exponential of a coefficient can be viewed as an odds ratio for the likelihood of being alive at the time of interest adjusted for other covariables included in the model (similar to a regular logistic regression[26]). So, in order to explore the effect of the abovementioned factors, they were added to the GRp models. This enabled us to explore their effect while adjusting for the other, patient and transplant-related confounders that were already found to influence survival.

Link between centre activity volume and JACIE accreditation time and link between type of transplant performed by the centre and centre activity volume

In order to evaluate the relationship between centre volume activity volume and JACIE accreditation time, Spearman correlation was performed. We decided to use this nonparametric measure due to the high proportion of transplants done in a non-accredited centre.

Moods median test was used to compare centre activity volume according to type of transplant. This test was performed on centre level.