Comparison of autologous and allogeneic hematopoietic cell transplantation strategies in patients with primary plasma cell leukemia, with dynamic prediction modeling

Sarah Lawless,[1*] Simona Iacobelli,[2*] Nina Simone Knelange,[3] Patrice Chevallier,[4] Didier Blaise,[5] Noel Milpied,[6] Roberto Foà,[7] Jan J. Cornelissen,[8] Bruno Lioure,[9] Ruben Benjamin,[10] Xavier Poiré,[11] Monique C. Minnema,[12] Matthew Collin,[13] Stig Lenhoff,[14] John A. Snowden,[15] Stella Santarone,[16] Keith M. O. Wilson,[17] Fernanda Trigo,[18] Peter Dreger,[19] Lara H. Böhmer,[20] Hein Putter,[21] Laurent Garderet,[22] Nicolaus Kröger,[23] Ibrahim Yaukoub-Agha,[24] Stefan Schönland[25] and Curly Morris[26]

[1]Belfast City Hospital, Belfast, Northern Ireland, [2]Tor Vergata University, Rome, Italy, [3]EBMT Data Office Leiden, Leiden, the Netherlands, [4]CHU Nantes, Nantes, France, [5]ICRCM, INSERM, CNRS, AMU and Institut Paoli Calmettes, Marseille, France, [6]CHU Bordeaux, Pessac, France, [7]Università La Sapienza`, Rome, Italy, [8]Erasmus MC Cancer Institute, Rotterdam, the Netherlands, [9]Nouvel Hopital Civil, Strasbourg, France, [10]King's College Hospital, London, UK, [11]Cliniques Universitaires St. Luc, Brussels, Belgium, [12]University Medical Center, Utrecht, Utrecht, the Netherlands, [13]Freeman Hospital, Newcastle, UK, [14]Skanes University Hospital, Lund, Sweden, [15]Sheffield Teaching Hospitals Foundation Trust, Sheffield, UK, [16]Ospedale Civile, Pescara, Italy, [17]University Hospital of Wales, Cardiff, UK, [18]Hospital Sao Joao, Porto, Portugal, [19]University of Heidelberg, Heidelberg, Germany, [20]Haga Teaching Hospital The Hague, the Netherlands, [21]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands, [22]Hospital Saint Antoine, Paris, France, [23]University Hospital Eppendorf, Hamburg, Germany, [24]CHU de Lille, University of Lille, INSERM U1286, Lille, France, [25]University of Heidelberg, Heidelberg, Germany and [26]Queen's University of Belfast, Belfast, Northern Ireland

*SL and SI contributed equally as co-first-authors.

Correspondence: S. Schönland
Stefan.Schoenland@med.uni-heidelberg.de

Received: December 21, 2021.
Accepted: May 27, 2022.
Early view: June 30, 2022.

https://doi.org/10.3324/haematol.2021.280568

Supplementary Material for the Comparison by transplant strategy

1. Methodology: Time bias and possible approaches

2. Table-Graft versus Host Disease and Response

3. Landmark analysis (Methods and Results)

4. Dynamic prediction method

1. Methodology: “Immortal” time bias and possible approaches

Time bias is likely to affect analyses comparing treatment strategies given in two or more steps in non-interventional studies, as in this one considering tandem Auto-Allo and tandem second Auto. The problem arises as groups cannot be defined and compared as if they were known at time 0 (here, the day of first transplant). For example, in this study at time 0 it is not known whether a patient who got Auto as first transplant will receive a tandem second Auto or a tandem Allo, or remain a “Single Auto” case. Importantly, in order to receive a tandem second transplant, this patient must survive relapse-free during the first months after first transplant; ignoring this “waiting time” and classifying cases from time 0 (by using information from their follow-up records) would systematically include the cases who fail early (or, too early to receive the second transplant) into the “Single auto” group, associating it to poor outcome by construction.

Immortal time bias is often overcome by assessing the differences between treatment strategies in a Cox model with time-dependent covariates; this was done in our study (results shown in Table 4). The main limitation is that the differences are thus evaluated as hazard ratios, while the associated survival probabilities are also of clinical relevance.

A simple way to show survival curves (or cumulative incidence curves) in this situation is to choose a “landmark” time when classifying the patients according to the treatment received up to that time and starting the comparison of outcomes. The results of this analysis in our study are reported in section S3. The landmark analysis has clearly a limitation in that it is affected by the choice of the landmark time (LT), which is in general arbitrary. Picking an early LT can leave a large proportion of patients not yet classified in the correct group, for example in our study at LT=1month most of the patients have received only the first auto transplant: the two groups of the tandem strategies are very small, and the Single Auto group is an heterogeneous collection of cases with many who later will receive the second transplant. On the other hand, a late LT implies a strong case selection, as patients failed before LT are excluded from the comparison. In our study we fixed LT=4month being close to the median time to second transplant, but in particular the resulting Allo-as-First group is heavily selected, including only the patients who survived the high risk of death of the first 100 days post allogeneic transplantation. However, this problem would occur even with a different choice of the LT time, making the use of landmark analysis particularly unsatisfactory in our study.

More complex statistical methods to estimate survival curves from Cox models with time-dependent covariates in presence of treatment strategies given in two or more steps are multi-state modelling [19] and dynamic prediction by landmarking [20]. The latter was applied in our study (results shown in Figures 2 and 3) and it is further illustrated in section S4.

2. Graft versus Host Disease

Table S1. Graft versus Host Disease

		Tandem Auto-Allo	**Allo-first**
Acute GvHD*	No aGvHD	56 (47.9%)	32 (48.5%)
	Grade I	26 (22.2%)	14 (21.2%)
	Grade II	23 (19.7%)	10 (15.2%)
	Grade III	6 (5.1%)	6 (9.1%)
	Grade IV	4 (3.4%)	4 (6.1%)
Chronic GvHD°	Cum. Inc. At 36 mo	56.2% (45.4, 67.0)	41.6% (26.8, 56.3)
	Cum. Inc. At 60 mo	58.1% (47.2, 69.1)	54.7% (39.1,70.3)
	% Extensive cGvHD	45%	64%

*Acute GvHD: number of cases and %. Percentages computed among non-missing cases. AGvHD information missing in 3 (2.5%) and 4 (5.7%) cases respectively.

° Chronic GvHD: cumulative incidence estimates at different time points, with 95% confidence interval. Competing events: death and relapse or progression. N=30 (15.6%) cases could not be evaluated due to missing info (19, 15.6%, and 11, 15.7% respectively in the two groups). The % of Extensive cGvHD is computed among all cases who experienced cGvHD.

Table S2. Response post-transplant

Transplant strategy			Frequency	Percent
Single Auto	Valid	CR	214	54.7
		VGPR/PR	158	40.4
		MR/SD	5	1.3
		Rel/Prog	14	3.6
		Total	391	100.0
	Missing	NA/NE	51	11.5
	Total		442	
Tandem Auto-Allo	Valid	CR	57	48.3
		VGPR/PR	58	49.2
		MR/SD	3	2.5
		Total	118	100.0
	Missing	NA/NE°	4	3.3
	Total		122	
Tandem Auto-Auto	Valid	CR	50	43.5
		VGPR/PR	64	55.7
		MR/SD	1	0.9
		Total	115	100.0
	Missing	NA/NE°	2	1.7
	Total		117	
Allo-as-First	Valid	CR	35	62.5
		VGPR/PR	18	32.1
		Rel/Prog	3	5.4
		Total	56	100.0
	Missing	NA/NE	14	20.0
	Total		70	

The % of Not Available / Not Evaluable is computed over the total of the group. The % of CR, VGPR or PR, Minimal Response or Stable disease, and of Relapse/Progression are computed over the total of cases available in the group.

°Based on information collected at 2nd transplant, we know that the 4 missing in the Tandem Auto-Allo group and the 2 missing in the Tandem Auto-Auto group had either CR or VGPR or PR.

3. Landmark analysis

Statistical methods

The approach and its limitations were introduced in the section S1. The landmark time LT was 4 months; for each endpoint (OS, PFS, CIR and NRM) the number of cases evaluable (alive event-free at 4mo) and the distribution according to the treatment received up to LT are reported in the tables. Unadjusted analyses were based on Kaplan-Meier probability estimates and Log-Rank test for OS and PFS, and on crude cumulative incidence and Gray test for CIR and NRM (Figure S1). Table S3 reports outcome estimates at time 60mo from 1st transplant with 95%CI limits and test p-values. Adjusted analysis was based on Cox models. Because of the strong bias affecting the Allo-first group, unadjusted tests were repeated and the Cox models were applied excluding this group. In the models, the baseline treatment group is Single Auto (Table S4).

Results

Table S3. Landmark analysis. Unadjusted.

	OS (N=663)		PFS (N=612*)		CIR	NRM
	N	estimate at 60mo (95%CI)	N*	estimate at 60mo (95%CI)	estimate at 60mo (95%CI)	estimate at 60mo (95%CI)
Single Auto	449	32.1% (26.7-37.5)	404	13.4% (9.2-17.5)	79.8% (75.1-84.5)	6.8% (4.0-9-6)
Tandem Auto-Allo	84	38.7% (25.6-51.8)	84	30.8% (19.4-42.2)	59.2% (47.2-71.2)	10.0% (3.4-16.5)
Tandem Auto-Auto	77	29.4% (15.8-42.9)	76	15.0% (4.5-25.5)	79.4% (67.0-90.9)	5.6% (0.0-11.8)
Allo-first	53	41.2% (26.3-56.1)	48	25.9% (12.0-40.0)	48.9% (33.5-64.3)	25.1% (12.1-38.1)
p-value (excluding Allo-1st)		0.591 (0.525)		0.309 (0.353)	0.002 (0.073)	0.001 (0.244)

*Same sample size for CIR and NRM

Table S4. Landmark analysis. Adjusted.

	OS	PFS	CIR	NRM
Single Auto	1	1	1	1
Tandem Auto-Allo	0.85 (0.62-1.19)	0.83 (0.62-1.11)	0.77 (0.56-1.05)	1.49 (0.67-3.30)
Tandem Auto-Auto	1.00 (0.72-1.38)	0.99 (0.73-1.33)	1.01 (0.74-1.38)	0.71 (0.21-2.37)

Effects expressed as HR (with 95%CI) versus Single Auto as baseline. Cases of the 1st trx Allo excluded. Adjustment factors: Age and Disease Status at first transplant (not shown).

Figure S1 (a). Landmark OS curves.

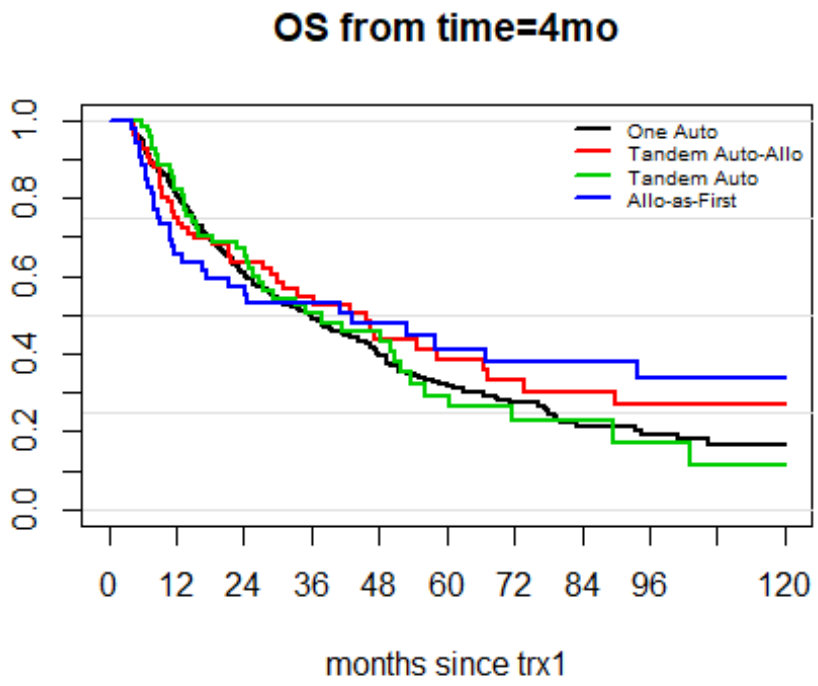


Figure S1 (b). Landmark PFS curves.

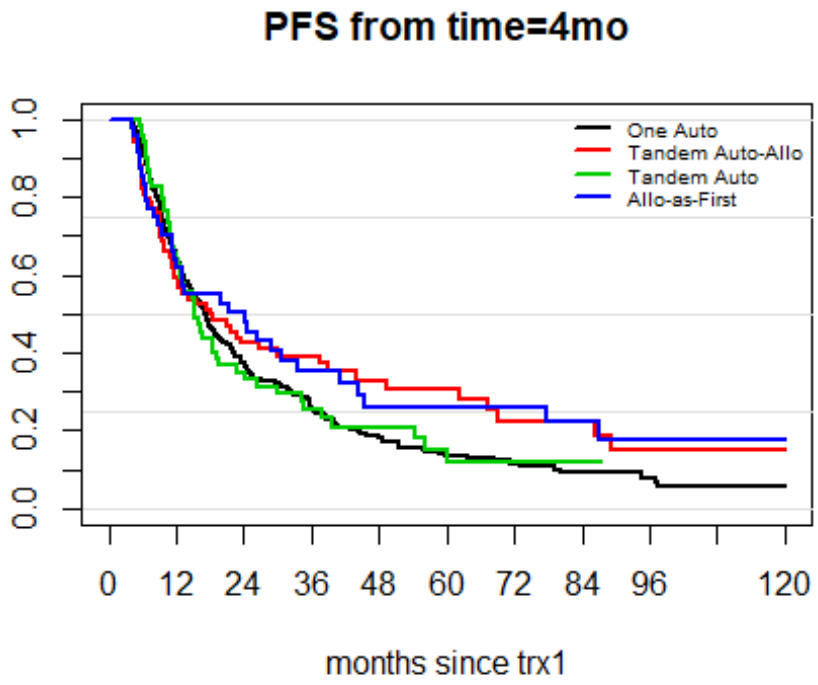


Figure S1 (c). Landmark CIR curves.

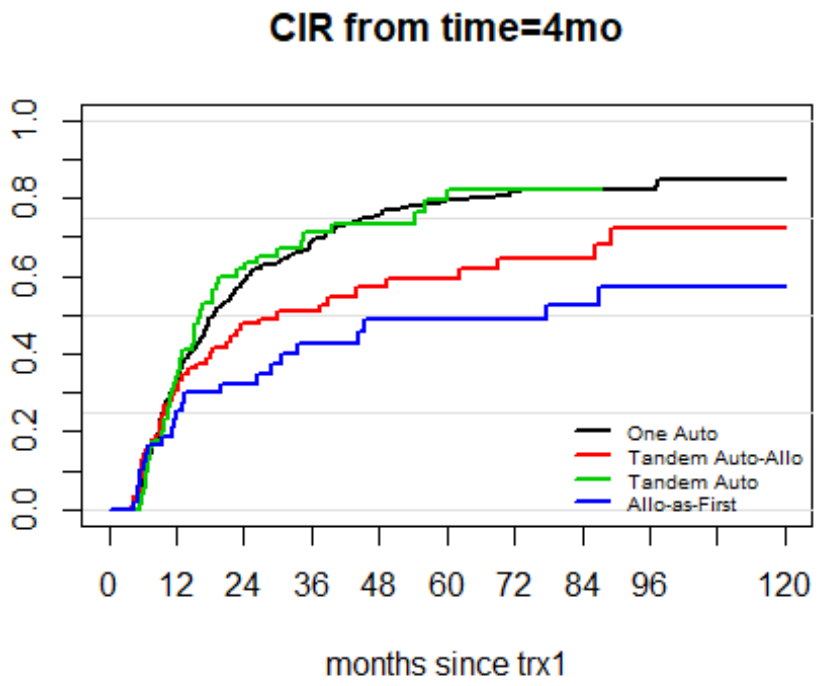
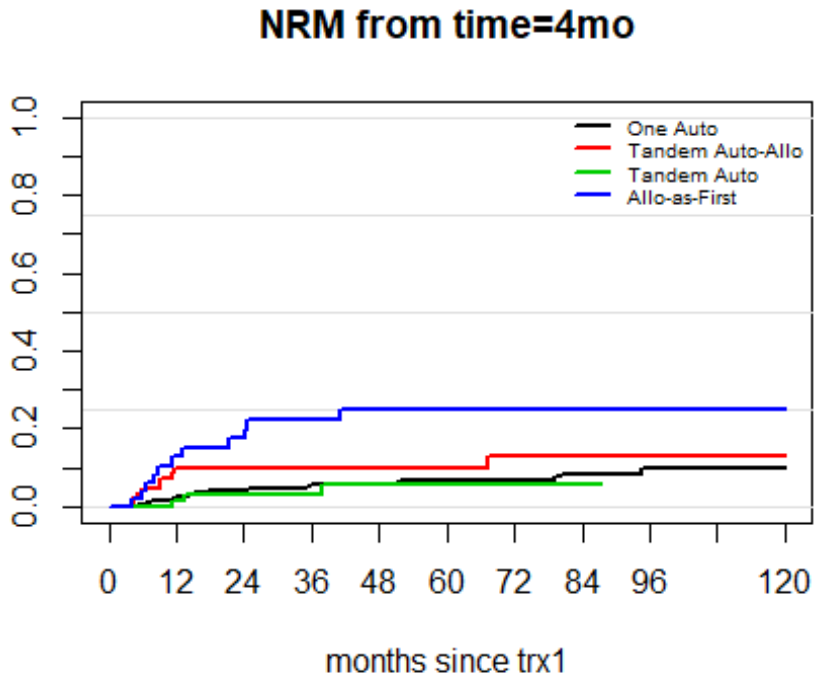


Figure S1 (d). Landmark NRM curves.

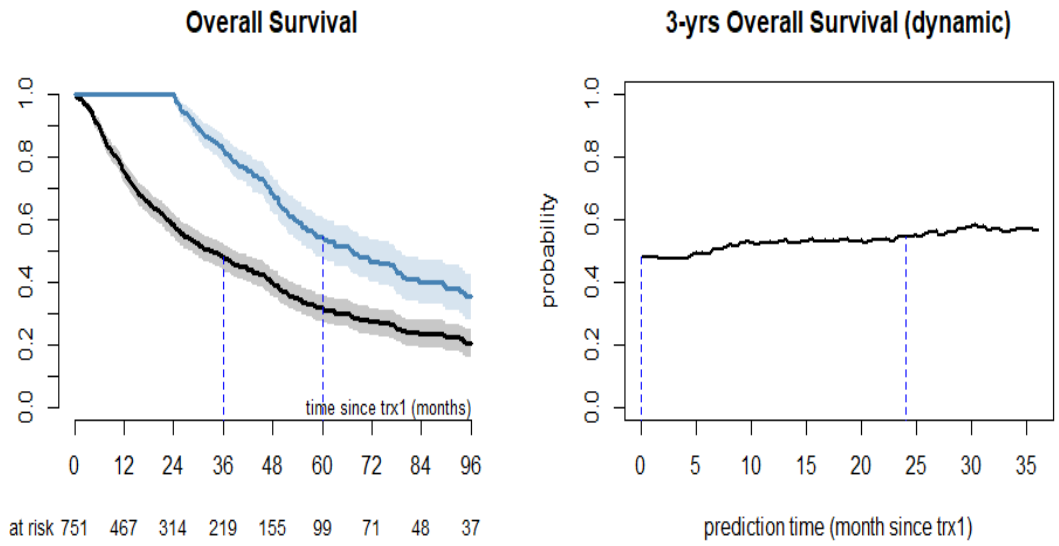


4. Dynamic prediction

The landmark analysis is the analysis of conditional probabilities for the patients who are still failure-free at the landmark time LT. It is interesting in itself, but limited by the choice of LT. Ideally, LT should be varied along an interval, say from t_0 to t_P, to appraise how the survival probabilities change according to the course of the disease. For example in our study moving the prediction time LT would allow to classify more and more patients with first transplant autologous into the groups of tandem Auto-Allo and Auto-Auto, and thus to evaluate the impact of these treatments. This is the intuitive principle of the dynamic predictions obtained by the method of “landmarking” (van Houwelingen, Putter[20]).

In this approach the focus is on estimating the survival probability after a certain “horizon” time since LT. In our study we considered of interest the probability of 3-yrs OS and of 1-yr PFS. We estimated these predicted probabilities moving LT from t_0=0 (the day of first transplant) to t_P=36mo. The graph below illustrates this concept using the Kaplan-Meier estimator. The method proposed by van Houwelingen and Putter estimates the dynamic prediction values from a “supermodel” which in intuitive terms combines the different landmark Cox models for each LT time.

Figure S2. Illustration of dynamic prediction curves.



Left panel: The black curve is the standard OS Kaplan-Meier curve estimated at time t_0=0 for all 751 patients included in the study. The blue curve is the landmark OS Kaplan-Meier curve estimated at time LT=24mo for the 314 patients still alive by that time. The focus is on the probability of surviving for 3 years after the prediction time, which on the black curve it is the value corresponding to time=36 (48%), and on the blue curve it is the value corresponding to time=60 (=24+36) (55%). Right Panel: These two probability values are reported on the curve for prediction time t_0=0 and LT=24mo respectively. The dynamic prediction curve joins the predicted 3-yrs OS probabilities from a number of landmark curves, showing the improvement of the 3-yrs OS for the PCL patients surviving during the first 36mo from first transplant.

In detail, we based our dynamic prediction estimates on 121 different landmark times. The supermodel was stratified on LT. The analysis was performed in R v. 3.5 using the library “dynpred”.